Aalto University
School of Business

The rise of passive investing, its effects on the underlying securities and implications on market efficiency

Master's Thesis

Oskari Eklund
Aalto University School of Business
Finance

Fall 2018

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Oskari Eklund

Title of thesis The rise of passive investing, its effects on the underlying securities and implications on market efficiency

Degree Master of Science in Economics and Business Administration

Degree programme Finance

Thesis advisor(s) Peter Nyberg

Year of approval 2018 **Number of pages** 66 **Language** English

Abstract

Over the last couple of decades, stocks have become increasingly owned by passive methods of investing, a type of investing that mechanically follows stock indices, and where decision-making isn't based on any research. This poses an interesting question about the pricing and efficiency of individual stocks that have had pressure from passive non-fundamental trading.

Therefore, in this master's thesis, I form time-series of passivity for individual stocks by aggregating the ownership of passive vehicles (index funds and exchange-traded funds) and analyse whether the quarterly flows and level of passivity alter the returns characteristics of stocks.

I find a statistically significant negative, though small, effect of passivity on the realized quarterly skewness, and further analysis indicates that it is due to an elevated realized negative deviation, consistent with the idea that increases in passive ownership are associated with higher reversals due to passivity's initial positive effect on prices and subsequent corrections induced by active investors.

Additionally, I study the cross-section of stocks by running the same regressions by group, formed from the level of realized return on equity (ROE) and find that passivity's effects are more prominent for stocks that have worse fundamentals, implying that the added non-fundamental buying pressure of passive institutions may have larger effects for stocks that wouldn't perhaps otherwise be in much demand by investors, and have hence a higher share of non-fundamental trading out of the total trading.

Finally, I study the potential long-term effects of passivity by forming event studies around index-deletions and find that the lagged level of passive ownership is a significant factor in explaining the abnormal returns around the deletions. Interestingly, the explanatory power of passive ownership is most prominent especially in the window leading up to the announcement of the index-deletion, implying that active investors may be more willing to take larger anticipatory bets against stocks that have had more non-fundamental flows by passive institutions in the past, consistent with the notion that each passive dollar may be adding something extra to the valuation levels of stocks and that arbitrageurs are limited in their capabilities in fixing this in the short-term.

This study most closely relates to previous literature by Qin and Singal (2015), Coles et al. (2018), Baltussen et al. (2017) and Israeli et al. (2017).

Keywords passive investing, market efficiency, skewness, negative deviation, event study, index-deletion -effects

Aalto-yliopisto
Kauppakorkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Maisterintutkinnon tutkielman tiivistelmä

Tekijä Oskari Eklund

Työn nimi The rise of passive investing, its effects on the underlying securities and implications on market efficiency

Tutkinto Kauppatieteiden maisteri

Koulutusohjelma Rahoitus

Työn ohjaaja(t) Peter Nyberg

Hyväksymisvuosi 2018 **Sivumäärä** 66 **Kieli** Englanti

Tiivistelmä

Osakkeiden omistusrakenne on kokenut suuria muutoksia viime vuosikymmeninä, kun passiiviset sijoitusmuodot ovat kasvattaneet omistustaan. Passiivisen sijoittamisen ominaisuuksiin kuuluu sen mekaanisuus ja se, että sijoituspäätökset eivät perustu analyysiin, jolloin herää mielenkiintoinen kysymys osakemarkkinoiden hinnoittelusta ja tehokkuudesta sellaisten osakkeiden kohdalla jotka ovat saaneet paljon painetta passiivisesta sijoittamisesta joka ei perustu fundamentteihin.

Tässä maisterin tutkielmassa muodostan passiivisuuden aikasarjoja yksittäisille osakkeille, aggregoimalla passiivisten sijoitusmuotojen (indeksirahastot ja ETF:t) omistuksia kvartaalitasolla, ja tutkin miten passiivisuuden muutos ja taso vaikuttavat osakkeiden tuottojen käyttäytymiseen.

Tutkielmassa löydän passiivisuudelle tilastollisesti merkitsevän negatiivisen, joskin pienen, vaikutuksen realisoituneeseen kuukausittaiseen tuottojakauman vinoumaan, ja tarkempi analyysi indikoi tämän johtuvan merkittävämmästä realisoituneesta negatiivisesta varianssista. Tulokset ovat johdonmukaisia ajatukseen siitä, että passiivisuus johtaa suurempiin yksittäisiin negatiivisiin tuottoihin, johtuen siitä että passiivisuuden lisääntyminen aiheuttaa aluksi positiivista painetta osakkeisiin, jonka aktiiviset sijoittajat kuitenkin myöhemmin korjaavat.

Tutkin myös osakkeita poikkileikkauksin, tekemällä samat testit eri ryhmille, jotka perustuvat viimeisimpään oman pääoman tuotto (ROE) -arvoon ja löydän että passiivisuuden vaikutukset ovat vahvemmat osakkeille joilla on heikommat fundamentit. Tulos implikoi, että passiivisten instituutioiden aikaansaama ei-fundamentaalinen ostopaine aiheuttaa vahvempia vaikutuksia osakkeille joille ei mahdollisesti olisi muuten niin paljon kysyntää ja joilla siten kaupankäynti koostuu suuremmalta osin ei-fundamentaalisesta kaupankäynnistä.

Lopuksi tutkin passiivisuuden potentiaalisia pitkän aikavälin vaikutuksia muodostamalla tapahtumatutkimuksia indeksistäpoistojen yhteydessä. Tutkimuksissa löydän, että passiivisuuden viivästetty omistus on merkitsevä selittävä tekijä selittämään epänormaaleja tuottoja indeksistäpoistoissa. Mielenkiintoinen havainto on erityisesti se, että passiivisuuden merkitsevyys on erityisen vahva aikaikkunassa ennen tulevan indeksipoiston ilmoittamista markkinoille, mikä implikoi sitä, että aktiiviset sijoittajat ovat halukkaampia tekemään suurempia ennustavia sijoituksia sellaisia osakkeita vastaan, joilla on aikaisemmin ollut enemmän passiivisten instituutioiden aihettamaa ei-fundamentaalista ostopainetta. Tämä havainto on konsistentti sen ajatuksen kanssa, että jokainen passiivinen dollari saattaa lisätä jotain ekstraa osakkeiden arvostustasoihin ja että aktiivinen raha ja arbitraasi on rajoitettua korjaamaan nämä vaikutukset lyhyellä aikavälillä.

Tutkielma liittyy läheisimmin seuraaviin aikaisemmin julkaistuihin tutkimuksiin: Qin ja Singal (2015), Coles et al. (2018), Baltussen et al. (2017) sekä Israeli et al. (2017).

Avainsanat passiivinen sijoittaminen, markkinatehokkuus, jakauman vinouma, negatiivinen hajonta, tapahtumatutkimus, indeksipoistovaikutus

Table of Contents

List of Figures

List of Tables

I – Introduction

Passive investing is a megatrend of the stock markets. Since 2007 the AUM of passive equity vehicles has grown globally by 230% to over 6 trillion dollars[1] and is expected to climb even more in the future with projections showing that they will overtake actively managed assets no later than 2024[2]. The main reasons behind this shift has been the introduction of cheap and simple index tracking exchange traded funds (ETFs) and mutual funds, favourable regulation and enhanced technology[3]. The growth has not only come from new capital, but also at the expense of active funds, i.e. funds that purposely deviate from their benchmark indices in the search of excess returns, or alpha. From a theoretical standpoint, this move has been sensible as it is a pure mathematical fact that active investing is a zero-sum game and after fees, which are typically high, it is a negative sum game (Sharpe, 1991). So, holding all else constant, the average investor is usually better off investing in passive funds over time. Purchases by passive institutions don't rely on any firm-specific fundamental analysis, however, which is why in excessive quantities it is possible that their actions can decrease the accuracy of stock prices.

Active money, or the so called "smart money", on the other hand has an extremely important function in the capital markets: they use large amounts of resources on research to estimate the real, fundamental values of stocks and act on this gained knowledge. In doing so they push the prices of stocks closer towards their fundamental values, add liquidity and bring markets closer to efficiency, i.e. a situation where the prices of stocks reflect all the relevant information in the markets but nothing more (Fama, 1970). Perfect efficiency of prices is arguably impossible though (Grossman and Stigliz, 1980), and history shows that there have been situations where active investors have been unable to bring efficiency due to so called limits on arbitrage, or somewhat paradoxically they have been uninformed and thus unable to push prices back to their fundamental values. These notions raise an important research question regarding the short and long-term effects caused by non-fundamental flows of passive investors.

Research around the growth of passivity is still rather young and to the best of my knowledge there are two notable papers, by Qin and Singal (2015) and Coles et al. (2018), where the actual amount and the share of passive money aggregated from ETF and mutual fund ownerships is used to study the effects of passivity. The results from both of these papers are supportive of

[1] Financial Times 05/16: https://www.ft.com/content/2552ce62-2400-11e6-aa98-db1e01fabc0c
[2] Moody's 02/17
[3] Credit Suisse 01/17

decreased weak-form price efficiency as a result of passivity but they disagree in terms of the effects on semi-strong efficiency, implying that further research around the topic is required.

Therefore, in this master's thesis, I form time-series of passivity for individual stocks by aggregating the ownership of passive vehicles (index funds and exchange-traded funds) and analyse whether the quarterly flows and level of passivity alter the returns characteristics of stocks.

I find a statistically significant negative, though small, effect of passivity on the realized quarterly idiosyncratic skewness, and further analysis indicates that it is due to an elevated realized idiosyncratic negative deviation, consistent with the idea that increases in passive ownership are associated with higher reversals due to passivity's initial positive effect on prices and subsequent corrections induced by active investors.

Additionally, I study the cross-section of stocks by running the same regressions by group, formed from the level of realized return on equity (ROE) and find that passivity's effects are more prominent for stocks that have worse fundamentals, implying that the added non-fundamental buying pressure of passive institutions may have larger effects for stocks that wouldn't perhaps otherwise be in much demand by investors, and have hence a higher share of non-fundamental trading out of the total trading conducted.

Finally, I study the potential long-term effects of passivity by forming event studies around index-deletions and find that the lagged level of passive ownership is a meaningful factor in explaining the abnormal returns around the deletions. Interestingly, the explanatory power of passive ownership is most prominent especially in the window leading up to the announcement of the index-deletion, implying that active investors may be more willing to take larger anticipatory bets against stocks that have had more non-fundamental flows by passive institutions in the past, consistent with the notion that each passive dollar may be adding something extra to the valuation levels of stocks and that arbitrageurs are limited in their capabilities in fixing this in the short-term.

The remainder of this paper is organized as follows: Section II further covers the previous literature and motives behind this thesis, in Section III is described the hypotheses to be tested, Section IV covers the methodology and data of the tests, Section V includes the results of the tests whereas further robustness is tested and limitations discussed in Section VI. Finally, in section VII, I discuss the concluding remarks and suggestions for future research.

II – Motives and previous literature

What is passive investing?

First it is essential to specify what passive investing really is and what the different forms of it are. The most significant form is index funds and ETFs, and I define them as vehicles that mechanically follow indices that are simply formed by market cap, equal weighting or other non-fundamental reason. That is, when investors invest in passive funds, the funds channel the investments to individual stocks based on the stocks' relative weight in the index that the fund is tracking. Therefore, passive investing takes no view on the valuation of individual stocks or how they are valued relative to other stocks, but rather accepts the current levels and their relative weightings. Prior to index funds, if investors wanted to own passive portfolios, they had to do it themselves by observing the stock valuations, forming portfolios and rebalancing at predetermined intervals, which of course requires high amounts of capital and is very expensive. Whereas passive investing through index funds is a cheap way to own diversified portfolios and doesn't require much capital, which makes it attractive and is perhaps the main reason behind the growth of indexing in recent years.

Additionally, there are individual investors and institutions such as pension funds or even central banks[4] that build portfolios in a very similar manner as passive index funds. On top of this, even some active funds may have large positions that could be classified as being passive which is to some extent confirmed by Wermers and Yao (2010) who show that active funds tend to own a lot of the same stocks with similar weights as passive funds. Furthermore, in recent years there has been significant amount of mechanical stock buybacks by companies themselves, some of which can be thought of as another form of added passive ownership and a non-fundamental pressure for prices[5].

Finally, there are other types of funds that mechanically track indices, but the indices themselves are formed from other criteria than purely market capitalization, such as value,

[4] For example, the Central Banks of Switzerland and Japan have been big net purchasers of equities for some time now, although they mostly invest via ETFs: https://www.forbes.com/sites/johnmauldin/2017/06/22/the-swiss-national-bank-owns-80-billion-in-us-stocks-heres-the-catch/#4e2924875362 & https://www.bloomberg.com/news/articles/2017-07-19/japan-bourse-head-turns-surprise-critic-of-kuroda-etf-purchases

[5] Buybacks also automatically increase the share of total shares owned by passive institutions because companies purchase the stocks from active investors.

growth momentum or volatility. These are the so called "factor" funds or "smart beta" funds that combine both passive and active investing, which is why they can't simply be classified to either group. Clearly the passivity classification for these other types of methods is rather difficult, but it nonetheless poses an interesting question about how much "pure" passivity there really is if we could accurately estimate it all.

Stock market efficiency

Studies about stock market efficiency are at the heart of academic finance and they are traditionally traced back to the paper by Fama (1970)[6] who proposed a framework for assessing the level of market efficiency and argued that the stock markets are highly efficient in the sense that no investor can consistently achieve positive risk-adjusted returns. This argument came to be called as "the efficient market hypothesis" (EMH) and it was followed by various papers where its sensibility was tested.

Many of these papers documented "anomalies" that were inconsistent with the EMH such as returns predictability through PE-ratios (Basu, 1977), size (Banz, 1981), excess volatility of stock prices inconsistent with efficient incorporation of new information (Shiller, 1981) return reversals due to overreaction and behavioural biases (DeBondt and Thaler, 1985) and various papers about stock market seasonality (eg. DeBondt and Thaler, 1987). Perhaps the most convincing argument against EMH was formulated by Grossman and Stiglitz (1980) who argued that it is in fact impossible to have perfectly efficient markets, because gathering information is costly and investors will not conduct research without an expected payoff. In other words, if prices perfectly reflected all the available relevant information at all times, investors couldn't gain abnormal returns by analysing firms and therefore no one would do it. On the other hand, if no one did any research, the markets couldn't be very accurate. Therefore, for the expected payoff to exist, there needs to be some inefficiencies in the market.

In the context of this paper, if perfect market efficiency applied, everyone should invest passively and be satisfied with the market returns, because consistent risk-adjusted outperformance would be impossible and being active wouldn't bring efficiency to the markets. But as per the argument by Grossman and Stiglitz (1980), perfect efficiency is impossible

[6] His paper was actually a review of the theory and evidence released up to that point, but he was the first to tie them together and formalize a comprehensive framework.

because we need research and incentives to research to get closer to efficient prices. Therefore, full passivity is an impossibility and one can question the sensibility of the current trend towards higher passivity and less research-based investing.

Why passivity matters?

Sharpe's (1991) insight was that the stock market simply consists of two groups: passive investors and active investors. The difference between these groups is that passive investors hold the market (index) with equal portfolios whereas active investors each individually deviate from this with different portfolio weights. However, the active investors must *collectively* match the portfolio of the passive group and thus each group earns the market return on aggregate. Within the active group, the investors' returns vary greatly due to unique weights: i.e. when one investor overweights a stock relative to a stock's weight in the passive portfolio, another investor or a group of investors must underweight it in a similar fashion, hence the characterization "zero-sum game". And because the active investors pay higher fees due to trading and various other costs, their return on aggregate always loses to the return of the passive investors (Figure 1).

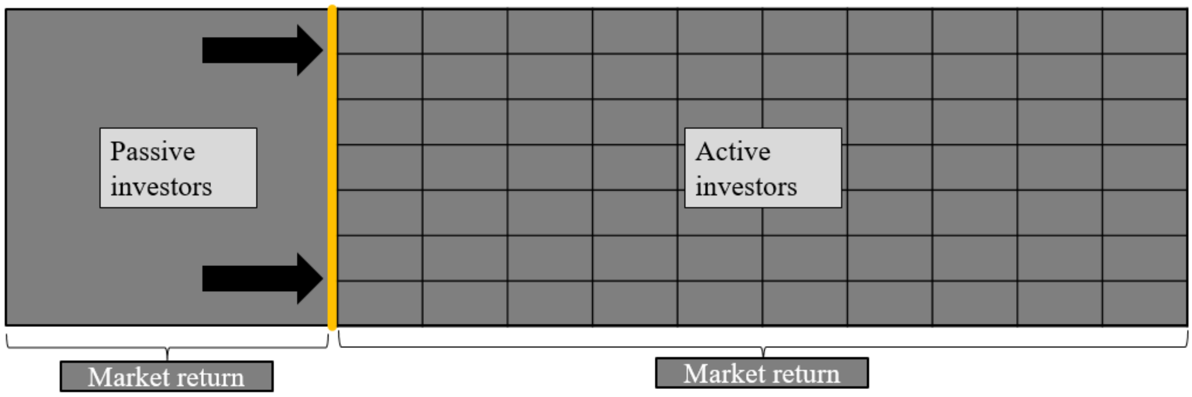


Figure 1: Sharpe's (1991) Arithmetic of active management visualized. The trend of passive flows can in this context alter market efficiency in two ways: 1. through the non-fundamental push of passive mechanical flows and 2. by altering the composition of the various active investors that would otherwise trade inefficiencies away.

In the simplified model depicted in Figure 1, if we accept that skill is involved in the returns of the active investors, i.e. the markets are not perfectly efficient and skilled investors can consistently pick good stocks over bad stocks, it is intuitive to think that over time the least skilled of the active group will exit the market completely or will switch to the passive group. This is because for the skilled to gain, the less skilled have to pay for it. Thus, when the average skill-level of the active group increases, it becomes increasingly difficult for active investors to outperform their peers and possibly even to find counterparties that are willing to trade stocks that they perceive to be mispriced. Therefore, the active part of the market requires sufficient enough amount of dispersion in the views of individual investors so that the most "skilled" investors can have the possibility to express their views about corporate valuations. In this context, the level and flows of passive ownership can alter market efficiency in two ways: first, by causing a non-fundamental push to the prices when they make investment decisions that aren't based on any research, and second, by altering the size and composition of the active group who normally corrects any deviations from fundamental values.

Furthermore, another intuitive way to think about passivity's potential adverse effects is to wonder what would happen at the limit, i.e. in a situation where stocks are completely or almost completely owned by passive institutions: first, new information wouldn't directly be incorporated to stock prices because passive institutions do not trade on fundamentals, they act only on investor subscriptions or redemptions, thus the valuations of fully passivized stocks would be fixed until perpetuity and hence they would in many ways mirror unlisted and untradeable stocks. This situation would likely be reached even in the case of "almost" full passivity, for active investors always require other active investors to trade with.

Clearly in these scenarios the stock markets would effectively stop working and this is not in any way sensible or possible which is why the cap for passivity must be somewhere significantly below 100%, as was similarly noted by Grossman and Stiglitz (1980). This notion also relates to the hypothesis by Pedersen (2018), who discusses how too high levels of passivity would lead to illiquidity in the secondary markets, increased cost of capital for firms and less informative stock prices. Throley (1999) also argued that if all investors were rational, they should only invest actively if they were in the top 25% of all investors because so many underperform the market over time. He then discusses that in a world of rational investors only the very best would stay active which would eventually lead to the diminishing of liquidity. His conclusion is that due to the overconfidence of many investors, they perceive themselves as belonging to the top 25% and therefore decide to stay active, which ultimately increases the liquidity of the markets and allows the truly skilled investors to enhance market efficiency.

Therefore, if we know for sure that full or very high amount of passivity will be bad for market efficiency, then one can hypothesize that getting closer to the limit may also be bad for efficiency (incremental deterioration). On the other hand, one could of course see passivity as a binary problem where we only observe the negative effects once a critical limit in passivity has been reached, i.e. when the amount of active investors is so low that trading becomes barely existent. Nevertheless, finding whether the negative effects of passivity on market efficiency are incremental or fully binary is an extremely important question for the future of the whole stock market itself to ensure efficient channelling of capital in the economy.

Previous literature

In theory, passive investing should only work as a follower of the markets, accept that the valuations driven by active investors are rational and hence have no price impact. However, many studies (Chang et al., 2015; Petajisto, 2011; Sullivan & Xiong, 2012) have shown how stocks that are added to an index experience a boost in their prices and increased co-movement with other constituents of the index, and how index deletions lead to opposite reactions. This is due to the fact that passive funds following the index are forced to buy the newly added stocks and sell the deleted ones because their goal is to simply track the performance of the index. Added co-movement comes from the fact that index funds trade the index constituents as a basket, meaning that for every additional dollar they receive from investors, they buy all (or a representative group) of the stocks within the index at their respective weights regardless of how the fundamentals for the companies look like.

The key insights from these papers are that the price movements resulting from being in an index are non-fundamental[7], i.e. it doesn't bring any new information and thus should have nothing to do with the valuation of the company[8] and that there is a downward-slope in the demand curve for stocks. Over time, if active investors are for some reason unable to reverse the valuation increases caused by passive money (they may for example fear that the flow of new passive capital will drive the mispricing even higher), the most highly indexed and "passivized" stocks can become overvalued. This is characterized by for example Wurgler

[7] Important notion in this context is that there are now actually more indices than stocks: https://www.bloomberg.com/news/articles/2017-05-12/there-are-now-more-indexes-than-stocks

[8] Although one could argue that being in an index can increase the liquidity for a stock as there are so many index funds trading them and that the added liquidity should warrant a premium.

(2010) as “detachment” and there are studies that show how this can already be a reality (Morck & Yang, 2001; Belasco et al., 2011).

The research on the effects of passive money is still very young and the amount of papers released is limited, but it is growing very quickly. The paper closest to this thesis is by Qin and Singal (2015) who find that stocks with higher passive ownership have stronger post-earnings announcement drift (PEAD) and deviate more from the random walk, implying decreased weak and semi-strong price efficiency. Later Coles et al. (2018) have also shown using regression discontinuity –methodology in Russell 1000 & 2000 -indices, that passivity is associated with a lower tendency for stocks to follow a random walk, but their results were contradictory to those of Qin and Singal (2015) in terms of the PEAD effects, and they concluded that the adverse effects from passivity are only apparent in the weak-form price efficiency. This contradictory finding gives room for further research around long-term pricing effects and abnormalities for heavily passivized stocks.

In other contexts, Ben David et al. (2014) find that stocks with higher ETF ownership have higher volatility, higher mean reversion and higher tail risk at times of market stress, suggesting that ETF inflows impound non-fundamental shocks to the underlying securities. Baltussen et al. (2017) find supporting evidence that the serial dependence (autocorrelation) of indices around the world has systematically turned negative and they argue that this is because high levels of indexing has started to drive prices which are then to some extent corrected by arbitrageurs.

Furthermore, Israeli et al. (2017) show that increased ETF ownership leads to higher trading costs for the underlying securities, increased co-movement among stocks, decreased predictive power of firm-level stock returns on future earnings[9] and a decrease in the amount of analysts following the firm, all supportive of declined pricing efficiency. Glosten et al. (2016) on the other hand find that for stocks in weak informational environments, ETF ownership actually somewhat increases the pricing efficiency as it correctly impounds systematic earnings information in to these stocks.

Passively managed funds have not only taken market share due to their lower fees and simplicity, but probably more importantly because their performance has been so good relative

[9] This is shown by Durnev et al. (2003): greater firm-specific volatility is associated with more informative stock prices.

to most active funds. In recent years, as much as 80-90% of active funds have lost to their benchmark indices regardless of fund type which is why many investors have likely switched to passive funds[10]. Arguably active fund managers possess a lot of skill and as per Sharpe (1991) they should on aggregate earn the same return as the passive funds, so it is an interesting question why so many perform poorly relative to the market. One reason that is also consistent with the detachment hypothesis is that due to the large inflows of passive money, there is a low frequency positive feedback loop in heavily indexed stocks that drives the performance of passive funds.

More accurately: each dollar invested in index funds is invested in the basket that they follow, pushing the underlying stocks higher and making the performance look good, which in turn attracts even more capital and allows the index funds to lower their fees[11], again leading to more capital invested and higher performance of the funds (Figure 2). And because it is the active investors from which the passive funds purchase the shares from, the relative performance difference can be attributed to the feedback loop of passivity[12]. If the hypothesized feedback loop of passivity continues going on, many active managers will keep suffering in performance versus the indices.

The loop described here is of course relevant only if there are impediments for arbitrageurs to reverse the effects on prices and that the demand curve for these stocks is steep enough that these passive demand shocks affect prices. In any case, the positive feedback loop is a rather plausible process through which the mispricing of the indexed stocks could come from and it relates to previous studies about fund flows as predictors of future performance (Warther, 1995; Coval & Stafford, 2007[13]; Staer, 2017) and returns chasing of investors (Sapp & Tiwari, 2004).

At the very least, it is rather intuitive to think that heavily indexed stocks are unlikely to be undervalued because the same limits on arbitrage wouldn’t apply, i.e. arbitrageurs would quickly buy the stocks because they would ultimately be backed by the seemingly ever-increasing passive money flowing in to the index funds. However, if the loop ever turned negative, i.e. systematic outflows from passive funds, this could be the case. This possibility

[10] Standard & Poors 12/16
[11] This is because they gather the fees as a % of the total AUM.
[12] In other words, passive institutions are gaining ownership from the high performing indexed stocks whereas active investors are losing the ownership, leading to a relative performance gap between the two groups.
[13] Coval and Stafford (2007) show how large inflows/outflows result in positive/negative pricing pressure for a fund’s existing positions.

was also recently noted by JP Morgan (2018)[14], and as a result the bank has begun advising their clients against investing in highly passivized stocks.

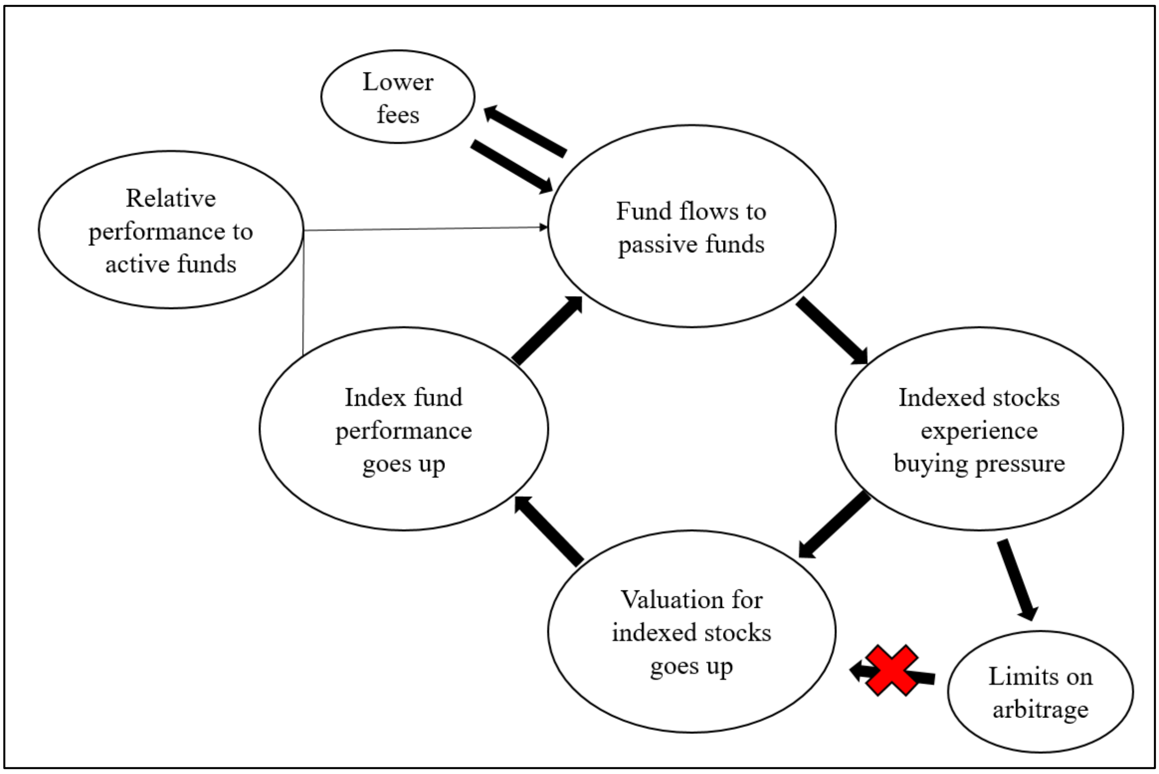


Figure 2: Positive feedback loop through which passive management can drive prices. Note that the process only works if there are sufficient limits to arbitrage.

Another topic regarding market efficiency is the idea about the loss of "the wisdom of the crowds" -effect which effectively means that as more and more investors pool their capital in to index funds, a lot of relevant latent information about the underlying stocks may be lost from the markets.

For example, The Bank of America frequently releases a report on their estimates on the share of active vs passive in the markets, and their latest report shows that their customers' cumulative net flows since 2007 to ETFs (thereby excluding mutual funds) has been around +\$200bn vs -\$350bn to direct stock purchases, implying that if direct trades by individuals include relevant information, this will likely disappear from the markets. The wisdom of the crowds may work well in many situations, but it is obviously quite difficult to test in the context of the stock

[14] Bloomberg 10/2018: https://www.bloomberg.com/news/articles/2018-10-24/jpmorgan-sees-7-4-trillion-passive-selling-pressure-in-downturn

markets, though Chen et al. (2014) find supporting evidence for the wisdom through social media -postings, showing how aggregate opinions can predict stock returns and earnings surprises.

However, contrary to the findings above, it could be argued that many or most of the investors who have gone passive are “noise traders”[15] who shouldn’t even be trading the stocks directly and thus the prices of the underlying stocks will actually be more efficient as the share of informed traders is higher in setting the prices. This would likely be true if the passive vehicles had no price impact, i.e. if they were truly just following the index. But in reality, as is shown in the research cited previously, there are price impacts which may be quite sticky through the increased co-movement and the feedback loop hypothesized in Figure 2, which is why the hypothesis about increased efficiency is somewhat problematic. Furthermore, as was discussed earlier, the fact that the average skill level of the active investors grows as the less skilled turn to passive funds may not necessarily be a good thing regarding pricing efficiency for it may lead to a lack of counterparties for the arbitrageurs when we get close to very high levels of passive ownership.

Before forming the hypotheses of this thesis, a more detailed look at the arbitrageurs and their limits is required. In the classical finance theory, arbitrageurs eliminate all pricing inefficiencies with little to no lag and can enter an infinite amount of arbitrage opportunities, eliminating the idiosyncratic risks they carry in the process. However, in practice arbitrage requires large amounts of capital which is usually externally sourced and thus under high scrutiny, it is risky and is usually conducted by specialized arbitrageurs that take large individual positions (Shleifer and Vishny, 1997; De Long et al., 1990).

Shleifer and Vishny (1997) also add that arbitrage is notoriously difficult in the equity markets as stocks are typically hard to value and the markets may stay inefficient even for long time-periods. This is supported by Wermers and Yao (2010) who find that high levels of passive mutual fund ownership in stocks leads to prolonged pricing anomalies which is evidence for active investors’ unwillingness to trade these stocks and Petajisto (2010) who analyses the NAV deviations of ETFs relative to their holdings and concludes that at least some of the difference comes from high limits on arbitrage. In this context, it is plausible to think how arbitrageurs might want to avoid trading against the most heavily indexed stocks because the trend has

[15] This is a popular characterization by Fisher Black (1986) for unskilled individual investors that trade purely on noise, i.e. market movements.

clearly been so strong towards passive investing. Nevertheless, the longer and the more prominent the potential pricing inefficiencies become, the better the expected payoff of trading against the inefficiencies become. Over very long periods the markets should thus always converge towards efficient prices, so long as there are sufficient amounts of research and capital willing to trade stocks.

Additionally, and somewhat unintuitively, even active managers may be incentivized to trade with the flow of passive money instead of against it: as is discussed by Baker et al. (2011), active managers benchmarked against a simple index are more likely to pick high beta stocks regardless of their potential alpha[16], because it yields better results in terms of the information ratio[17]. In other words, stocks that have received an artificial bump in their betas due to the increased co-movement are more favourable investments regardless of their fundamentals. Also, if there truly is a positive feedback loop of passivity as characterized in Figure 2 then if active investors are aware of it and understand that it warrants a premium to passivized stocks, then they may take positions purely as a result of the fact that there is non-fundamental demand for these stocks.

Noteworthy is however, that the lending supply required for shorting indexed stocks is likely rather large as index funds tend to be quite willing to lend their positions with relatively low fees and they are also a favourable counterparty for the borrowers in terms of recall risk (D'Avolio, 2002), implying that shorting these stocks shouldn't be that difficult.

[16] Potential alpha can be characterized as the perceived mispricing of a stock by an active manager.
[17] They discuss how maximizing Sharpe ratio (excess returns over risk-free rate) instead of information ratio (excess return over tracking error) would yield better results in terms of performance.

III – Hypotheses

If the detachment hypothesis discussed above is true and that the non-fundamental flows from passive investors affects prices, over time we should see at least some reversals in prices because arbitrageurs are not fully limited in their actions and the markets will converge towards efficiency over time. The reversals could be due to fundamental information being released that doesn't support the slow and steady non-fundamental push of passive money, hypothesized in Figure 2, and the idea is that once the market as a whole sees the potential pricing failure, the reversal is much more likely than when only hypothesized by smart money. I do not necessarily posit that the passivized stocks are inefficiently priced all the time, but rather that the way in which they approach efficiency is affected by the added passivity and its non-fundamental flows.

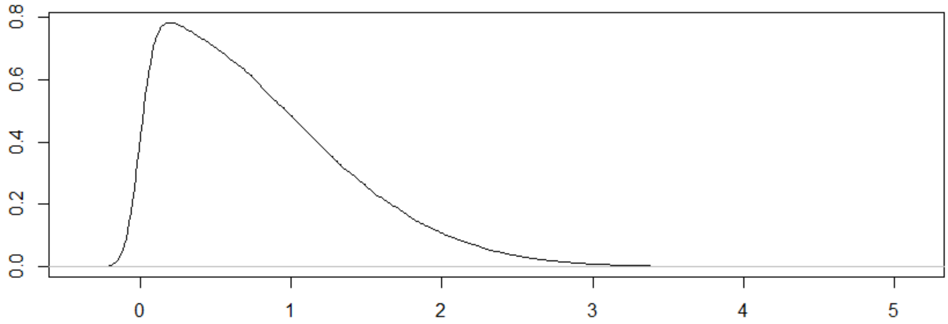
The first hypothesis of this thesis is related to the studies by Qin and Singal (2015) and Coles et al. (2018), and their finding about the diminished randomness of stock returns as a result of passivity. It further links to the papers by Baltussen et al. (2017) who hypothesized that the change in autocorrelation at the index-level is likely due to arbitrageurs correcting price impacts of index funds and Ben-David et al. (2017) who showed that ETFs as a product increases stock-level tail risks.

To build on these findings, I further study the shape of the realized returns distributions in the context of passive flows at the stock-level and hypothesize that:

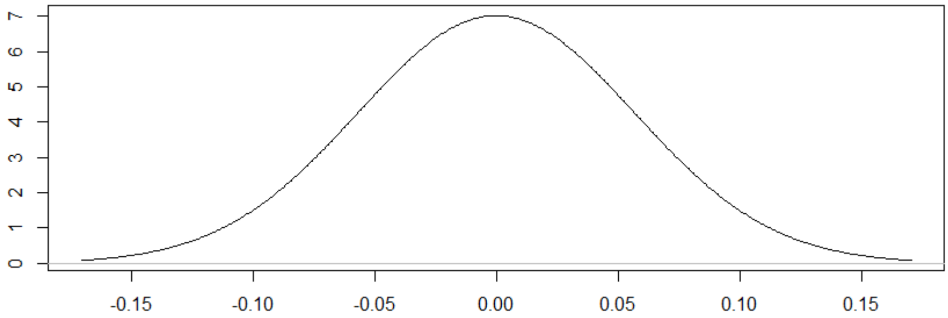
Hypothesis 1: Passivity increases the propensity of reversals for stocks.

To test if passive money has the non-fundamental push and then leads to the reversal effects described above, the stocks with larger passive flows should exhibit smaller or negative skewness in their returns distribution and that the effects on skewness should come from more prominent left-tail movements, i.e. an elevated negative deviation. In Figure 3 I have plotted simulated distributions where the concept of skewness is easily understandable: the longer the tail is to a direction the more skewed it is towards that direction. Hence, stocks with higher tendency for reversals should graphically look closer to Figure C rather than A or B. Stock returns can't, however, be less than -100% which is why the left tail in the distributions is capped to that value.

A: Positive skewness



B: Zero skewness



C: Negative skewness

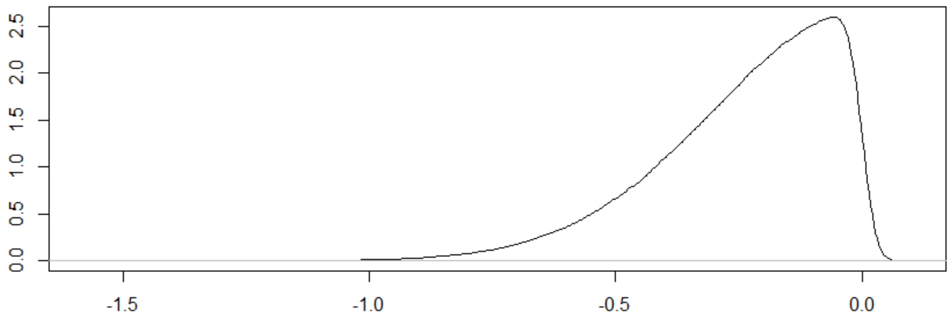


Figure 3: Simulated examples of distribution skewness

On the other hand, it is likely that for some stocks the added buying pressure is likely justified, whereas for others not so much. For example, firms that consistently release good earnings and are profitable would more likely be bought by investors in any case whereas firms that do poorly wouldn't probably have much of the buyers without the flows of passive institutions. These factors can be significant for the overall demand of a stock and I thus hypothesize that passivity's effects are not constant across the board:

Hypothesis 2: There are cross-sectional differences in the effects of passivity.

H1 and H2 are conducted to test the potential short-term effects caused by the increase of passive ownership and whether they are associated with reversals or not. However, it may be that due to the rather constant flow of new passive money entering the markets, these potential short-term reversals can't fully accommodate for the added buying pressure from passive vehicles. This may be the case if for example the selling pressure caused by active managers is always offset by new passive money entering index funds that are forced to buy even the stocks that are clearly mispriced in terms of their fundamentals. If the active managers recognize this ex-ante then they will avert entering a trade against highly indexed stocks which could further enhance the detachment effect. This is the previously mentioned limit to arbitrage, consistent with the hypothesized feedback loop, and it can be characterized as "noise trader risk by indexing". Additionally, even if there are reversals found as a result of passive money flowing in and affecting the prices, it is very much possible that some premium still persists. For example, simply the fact that a stock may be a target of future non-fundamental flows can be a factor that affects a stock's valuation. These premiums should, however, disappear quickly once a stock gets deleted from a prominent index because after a deletion it is no longer a target of future passive money. I thus hypothesize that:

Hypothesis 3: Passivity leads to strengthening of the index-deletion effects.

Index-deletion and addition -effects have been studied extensively before and the finding that stock deletions lead to high negative abnormal returns is shown in multiple papers (for example Petäjistö, 2011). But to my knowledge the potential drivers behind these effects have not been studied very thoroughly and the previous research tends to assume that every index-deletion is comparable. However, if we know that passive non-fundamental flows can affect the pricing process of stocks, then it is intuitive to look at the deletions in the context of passive ownership: stocks that have more passive ownership have had more passive flows and hence potentially a higher premium. The findings for this hypothesis would be consistent with Figure 2, i.e. a potential low frequency positive feedback loop driving the valuations of indexed stocks and due to persistent limits on arbitrage, active investors can't reverse the effect completely until the stock gets deleted from a notable index. Hence, a more proper characterization of the index-premium could perhaps be "the passivity premium".

IV – Methodology and data

Methods for H1: Realized shape

Skewness

To test the first hypothesis, I form returns distributions over time from daily returns and analyse the skewness of the distributions. I calculate the distributions from each stock's idiosyncratic returns, i.e. firm-specific returns and these are obtained by simply subtracting the return of the market[18] from the return of passive stock i at time t (Equation 1).

Firm-specific returns are used instead of pure returns because there is already strong evidence that stocks move differently once they are included in indices and have passive ownership. Thus, to purify this effect to at least some extent, broad market returns are removed from their returns to approximate firm-specific returns. Additionally, if there are indeed higher reversals for stocks that have had non-fundamental pressure, and that the reversals are more likely to occur when firms release new information that doesn't support the added pressure, then I would expect to see the effects more prominently in the firm-specific component of stock returns. The return calculation used in all calculations is the simple relative return.

Idiosyncratic returns of stock i at time t are calculated by:

$$IR_{it} = R_{it} - R_{jt} \tag{1}$$

There are also other ways in which one can estimate firm-specific returns, such as running stock-specific Fama-French, CAPM or other factor -regressions and using the residuals of these regressions as the firm-specific return. However, the way in which the returns are calculated is not a significant factor when calculating the measures of shape for a stock's returns distribution, for these measures mainly give weight to the most extreme movements of a stock. The most

[18] It is assumed that the return of the Russell 3000 -index is the market return. This should be an accurate market proxy for the index covers approximately 98% of the U.S. equity market.

extreme movements of stocks are mostly due to idiosyncratic movements[19] and as these are automatically given much more weight in the measures of distribution shape, one does not need to overly stress the way in which firm-specific returns are calculated. Evidence for this finding is provided in the paper of Chen et al. (2001) whose skewness forecasting -results are stable across multiple returns specifications.

If it truly is the case that passivized stocks have a higher propensity for left tail -movements, then we should see a link between smaller/negative skewness in their returns distributions and the ownership and flows by passive money. Skewness[20] is measured by the variable IRS (idiosyncratic return skewness):

$$IRS_{it} = \frac{\sum_{i=1}^{N} \frac{(IR_{it} - \overline{IR})^3}{N}}{s^3} \tag{2}$$

Where $\overline{IR}$ is the mean idiosyncratic return over period t and s is the standard deviation of IR over the same period. The intuition behind the skewness formula is clear: each realization is demeaned, raised to the third power, summed and then standardized for comparability. It thus gives higher weights to extreme realizations and tells us about the tails of a distribution and which tail tends to dominate the other. A highly negatively skewed distribution simply states that the most extreme movements tend to be negative.

I then run two types of regressions to test the effect of passive ownership and flows to realized skewness. First, in a standard OLS form such as below:

$$IRS_{it} = a + bPassivity_{it} + cZ + \varepsilon_{it} \tag{3}$$

Where the dependent variable is the measure of the realized skewness for stock i during quarter t. I analyse quarterly changes due to the fact that a more frequent time-period wouldn’t be possible because data limitations and a more infrequent period would include multiple

[19] For example, if a stock moves up 10% in a day, the finding is automatically given much more weight in the skewness formula, than a day on which the return was 1%. As the market or other factors rarely move up 10% in a day, the finding is likely mostly due to firm-specific movements. The same intuition applies for negative returns.

[20] This is the Fisher-Pearson measure for skewness.

observations of many of the key variables. Passivity represents the variables of interest, i.e. passive ownership and its change or lagged change and Z represents a group of control variables that also affect the shape of a stocks returns distribution. This initial regression setting is similar to that of Ben-David et al. (2017) who studied how ETFs as a product can affect stock-level skewness, but differs importantly in two aspects: 1. I study the combined effects of all passive non-fundamental flows and 2. I use the firm-specific component of stock returns as was discussed earlier.

The variables of most interest are the quarterly passive ownership-variables. First, the ownership share of passive money for individual stocks ($P\_Own_{\,it}$), calculated by simply dividing the aggregate shares owned by passive institutions with the total amount of shares outstanding. Second, related to this variable is the flow of passive money ($\Delta P\_Own_{\,it}$), i.e. the change of passive ownership, as that is really the factor that would be driving the valuations in the short-term. For example, a stock may not have had any new flows of passive money recently but it may still rank very highly in the share of passive money owning it. Hypothetically, these kinds of stocks would be accurately priced for mostly active money would have traded them more recently. The third important variable is the lagged change in ownership ($\Delta P\_Own\_Lagged_{\,it}$), i.e. the ownership change in the previous time-period, for it is sensible to think that the potential effects, especially the reversals would take some time to occur.

To control for market-specific factors driving the shape, I control for the realized skewness and standard deviation at the market-level (Russell 3000). Firm-specific control variables of interest are: P/E ratio, P/B ratio, P/S ratio, stock volume, number of analysts following the firm and the quarterly earnings surprise of a firm. The first three should capture the “glamour effect”, i.e. stocks that have high future expectations impounded in their prices and they may thus be riskier. Volume is added as in some extreme market moments a stock’s liquidity can diminish and thus lead to higher tail movements. Number of analysts following a firm should decrease tail movements, assuming that the markets follow the analysts and their estimates are accurate. EPS surprise controls for the fact that quarterly earnings are probably the single most important driver of extreme stock-specific movements each quarter.

Finally, I also control for the size of a firm, its lagged return, lagged stock volume, lagged standard deviation and lagged realized skewness because these are shown to predict skewness (Chen et al., 2001).

Interestingly, on average the skewness for individual stocks tends to be positive, whereas somewhat paradoxically, the skewness at the index level is usually negative. Stock-level positive skewness may be explained with the fact that stock prices tend to on average exhibit positive returns and on average have an equal probability of experiencing movements from either tail of their distribution with the notion that the left tail is capped at -100%, for stock prices can't be negative (Figure 3). The negative skewness of indices can be explained with the fact that indices are sort of like portfolios and due to diversification benefits, the overall volatility as well as large tail movements are less likely. But for a portfolio, the tendency to experience large negative returns is more probable than large positive returns because correlations tend to increase in periods of market stress (Albuquerque, 2012). So, the stocks within a portfolio have a higher tendency to co-move downwards rather than upwards and hence at the index-level, skewness is usually negative.

Even though the measures of returns skewness are so different at the index and stock -level, the index level skewness measure is nevertheless an important factor to control for, because it gives a good understanding about the market conditions during a period under analysis. Previous literature about forecasting and explaining skewness is very thin though, so it is possible that some meaningful variables are omitted from the analysis and this area offers meaningful topics for future research.

The second regression setting[21] with time and industry -specific factors is as follows:

$$IRS_{it} = a + bPassivity_{it} + cZ + Industry_i + Time_t + \varepsilon_{it} \tag{4}$$

Where IRS, Passivity and Z variables are the same as before. The differences in this setting are Industry and Time -variables that indicate industry and time -specific dummy variables, respectively. The time dummies control for the fact that each quarter is different from one another. This dummy should proxy a broader market state -variable than the mere Russell measures used earlier and considers factors that can't be controlled for explicitly. Industry dummies are used to control for the fact that the returns distributions of different types of firms

[21] This regression setting is more formally referred to as the "fixed-effects model"

are likely similar within groups and less so between groups. Ideally one would set a dummy for each different stock because each stock likely has some individual characteristics that drive their distributions, but due to the large amount of individual stocks in sample, it is not feasible in this study.

Negative deviation

Even if the results from the initial regressions would indicate that the effects of passivity on realized skewness are negative, it still doesn't fully answer the question of where the results come from and are there steep negative reversals involved with passive purchases. In other words, the result of decreased skewness can still imply that the skewness of a stock is positive albeit it is less positive than before. Negative effect on skewness would be consistent with Hypothesis 1, but it is also consistent with a scenario where stocks that have had large passive inflows have less large positive movements (i.e. smaller positive skewness) and they may or may not have a higher tendency to have larger reversals. Smaller positive skewness would actually be in line with previous research about the fact that heavily indexed stocks tend to co-move more together and exhibit smaller tail movements overall.

Therefore, to give a more comprehensive answer the question of potential reversals, I run a third regression where the distinction to the regressions before is that the dependent variable is now the idiosyncratic negative standard deviation ("downside deviation") of a stock:

$$N_{Dev_{it}} = a + bPassivity_{it} + cZ + Industry + Time + \varepsilon_{it} \tag{5}$$

Where N_{Dev} is calculated as:

$$N_{Dev} = \sqrt{\frac{\sum_{i=1}^{N}[\min(IR_t, 0)]^2}{N}} \tag{6}$$

The formula only considers the negative firm-specific returns and thus tells a more accurate story about the left-tail of a distribution, and a higher N_{Dev} implies that there are days on which the stock's return has been highly negative compared to normal negativity. Combined findings of passivity's negative effect on skewness and positive effect on negative deviation would strongly imply that the observed effects are due to larger reversals and not simply smaller tail movements overall.

Methods for H2: Cross-sectional differences of H1

The characteristic of special interest is the profitability of a company. The intuition here is that profitability is probably the most salient factor for a company and the ROE measure is the most widely used measure of a company's financial performance (Monteiro, 2006).

Furthermore, stocks with lower profitability wouldn't necessarily be bought by passive investors if they analysed the fundamentals of each investment they made, and hence the effects of passivity may differ according to the financial performance of firms. For example, if an active investor had to choose between two identical companies with differing profitability-metrics, the one with the better profitability would likely be the winner. In the case of passive investing, this analysis is omitted from the investment process, and the investor likely ends up owning both of the companies. In this scenario the decision to invest passively results in flows to both of the companies, whereas if the investor invested actively, only the better company of the two would have received the flows.

For the tests, I sort the sample stocks with the most recent ROE-measure at each quarter and run regressions 4 and 5 for these groups individually. ROE is calculated from the latest values released by dividing the trailing yearly net income by the total equity reported by the company. ROE is chosen as the main variable for these tests because it is calculated from the net income of a company which tends to be a key variable mostly observed by all market participants, it doesn't directly relate to the size of a company for it is standardized by the amount of equity capital and the variable is very easy to calculate and understand.

Methods for H3: Passivity on the index-deletion effects

To analyse the deletion effects, I employ a standard event-study methodology. I first calculate daily abnormal returns around the event periods similarly to Equation 1, where the abnormal return is the simple excess return over the Russell 3000 -index that represents the U.S. market return.

I then calculate the cumulative abnormal returns from the individual abnormal returns of each event for each day in the event window by:

$$CAR(t_1, t_2) = \sum_{t=t_1}^{t_2} AR_{i,t} \tag{7}$$

The cumulative abnormal returns across dates in the event window are then averaged to find the average cumulative abnormal returns at each time point:

$$CAAR(t_1, t_2) = \frac{1}{N} \sum_{t=t_1}^{t_2} CAR(t_1, t_2) \tag{8}$$

Where t_1 and t_2 indicate the start and end dates of the event window, respectively.

The statistical significance of the CARs/CAARs can simply be tested with the test statistic calculated by dividing the CAR/CAAR with its respective standard deviation.

Finally, the cumulative abnormal returns are further analysed through a simple regressions model similar to Equations 3 and 4, where CAR can be any chosen period around the tested event:

$$CAR_i = a + bPassivity\_lagged_i \ + \ cZ + \ \varepsilon_{it} \tag{9}$$

Data

Identifying passive funds and their holdings

I identified passive mutual funds and ETFs using a combination of databases and methodologies. Initially, I searched equity index funds from The Center for Research in Security Prices (CRSP) by selecting funds that had been flagged as being a domestic equity index fund or an ETF during my sample period of 2000 – 2017. However, using CRSP as the sole identification tool would have been problematic as CRSP doesn't make a clear distinction between different share classes of a fund. This problem is not apparent with Thomson Financial's CDA/Spectrum S12 -dataset (CDA), where each fund number truly indicates a unique fund. Thus, to solve for potential duplicates, I bridged the two databases using Mutual Fund Links (MFLINKS)[22]. I then searched by name from CRSP and CDA for remaining equity index funds that had not been flagged as being index funds or ETFs in the CRSP database, using methodology similar to that of Qin and Singal (2015).

Closet indexers were identified by first finding all domestic equity funds from CRSP that did not have the index fund/ETF flag, couldn't be specified as an index fund by name, had a high amount of individual stocks (>100) in holding, because tracking an index requires a large amount of individual stocks, had AUM of over \$50m to limit the sample size and had an average turnover of less than 40% per year, because index funds do not typically alter the compositions much year-on-year. The potential closet indexer -dataset was then merged with active share[23] dataset calculated by Martijn Cremers[24], where I was able to match 535 of 549 potential closet indexers using their MFLINKS identifiers. I then calculated the average active share from Cremers' dataset and flagged my initial dataset's funds as closet indexers if their average active share was less than 50% (as per. Petäjistö, 2009). Through my methodology I found 68 closet indexers between 2000 and 2017 which is a bit less than what has been estimated earlier. This is likely due to the multiple variables and stages that I used in the estimation.

---

[22] The CDA-MFLINKS-CRSP-structure was available through the Wharton Research Data Services (WRDS)

[23] Active share (AS) = $\frac{1}{2}\sum_{i=1}^{N}|w_{fund,i} - w_{index,i}|$, where w indicates the weight of the stock in index/fund i and N the total amount of stocks in the benchmark index.

[24] Cremers has calculated the yearly Active Share for numerous funds from 1990 – 2015. AS is calculated in the dataset by comparing holdings with multiple indices and the one with the smallest AS is chosen. This dataset is available at: https://activeshare.nd.edu/data/

I then re-used the dataset of Cremers to find all funds that had an average AS of less than 10% to find index funds that might not have been flagged by my earlier methods. These funds were also sanity checked by name. I also checked that the dataset didn't include any equal-weighted funds for their trading patterns could cause reverse causality in the regressions because they trade on portfolio movements (i.e. rebalance) and not simply investor subscriptions and redemtions (flows). Finally, I combined the funds from all methods, deleted duplicates and funds that Qin and Singal (2015) had excluded after analysing the prospectuses of each fund in their sample. The final passive fund dataset has 711 individual funds throughout the sample period. By my estimates, the amount of passive funds and assets over the period are somewhat larger than what Qin and Singal (2015), which is likely due to slightly different methods in searching by name and the multi-staged closet indexer -calculation that I used, but the estimates are well aligned overall. The amount of funds and their assets per year is shown in Table 1. The exact CDA fund codes are shown in Appendix A.

Table 1: Descriptive statistics on market passivity

Year-end assets under management for passive and active funds targeting the U.S. equity market. Passive fund AUM is aggregated from CDA's assets-variable. For years 2002, 2010-2013 the AUM is estimated from the reported holdings of the funds. For 2017 the AUM is calculated from the reported holdings in the CRSP database. For a more detailed discussion about the fund identification process, see Appendix A. U.S. Equity market cap is the total market cap -variable from CRSP.

Year	U.S. Equity Market Cap (billions $)	No. of Passive Funds	AUM of Passive Funds (billions $)	Total Passive Ownership %
2000	15 623,59	259	417,82	2,67 %
2001	13 845,40	311	408,61	2,95 %
2002*	11 027,06	329	283,56	2,57 %
2003	14 577,76	326	453,47	3,11 %
2004	16 450,54	358	647,35	3,94 %
2005	17 371,12	368	681,02	3,92 %
2006	19 607,35	357	787,98	4,02 %
2007	20 195,66	536	898,41	4,45 %
2008	12 129,15	590	915,45	7,55 %
2009	15 804,52	571	1095,72	6,93 %
2010*	18 490,58	536	980,56	5,30 %
2011*	17 886,83	513	1408,16	7,87 %
2012*	20 352,70	487	1575,86	7,74 %
2013*	26 281,33	470	2045,39	7,78 %
2014	28 964,03	463	2162,84	7,47 %
2015	27 658,78	444	2467,89	8,92 %
2016	30 150,55	436	2704,43	8,97 %
2017**	35 731,08	400	3359,60	9,40 %

* Estimated from reported holdings due to missing assets-data
** Estimated from CRSP holdings due to missing asset data.

After identifying the funds, I searched for their holdings from the CDA database where quarterly holdings data for mutual funds and ETFs is available. I then aggregated the holdings for each quarter using each stock's unique CUSIP-code and was able to calculate the aggregate ownership-% of each individual stock for each quarter-end by dividing the aggregated shares by CDAs shares out (SHROUT)-variable[25]. The amount of individual stocks in the sample is very high (N = 10 594) throughout the whole period. This is mainly due to the fact that some indices have thousands of stocks in them (most notably Russell 1000/2000 -indices). This is a good thing as it automatically brings a lot of variation to my sample and thus I do not need to

[25] More specifically, SHROUT *10 000

gather and identify a separate sample with low passive ownership -characteristics from another source. The CDA dataset for holdings was incomplete for some funds and quarters however as CDA doesn't list the smallest holdings[26] and my sample likely lacks some passive funds completely, at least non-US-based passive funds that have not been matched in the CDA-MFLINKS-CRSP-structure. It is thus likely that any estimates for passivity from my dataset have a downward bias.

The quarterly change in passive ownership (ΔP_Own) was calculated by simply subtracting the next quarter's observation's aggregated ownership-% by the previous one. Thus, the variable describes the % of a firm's total shares that have been bought/sold by passive funds during the quarter. For some stocks there were gaps in the ownership data and these were fixed by setting the next change observation as NA if the gap in holdings was more than three quarters. Otherwise this would have led to unwanted variation in the ΔP_Own-variable. However, if the gap was more than 2-years, it was assumed that that this was due to a stock's re-addition to an index and the change in ownership was set as the aggregate ownership of the first aggregated holdings observation after the long period. Clear outliers (if absolute change > 4%) were set as NA in the change variable, as these were most likely data points as a result of missing holdings data and would again lead to unwanted variation. These extremely large changes would be sensible only in cases where a stock was added/deleted from an index and some of these cases are thus lost from the sample.

Finally, instances where the change in holdings went from a large negative change in one quarter (< -1%) to a large positive one in the subsequent one (> + 1%) were likely a result of missing data in the negative change -quarter, and these changes were modified so that their values were multiplied by a factor of 0.25 so that the possible variation caused by missing data was minimized. These were not outright set as NAs because some of these swinging changes -instances have really been true. For example, in volatile market states the fund flows to mutual funds and ETFs may experience large swings from quarter to quarter. ΔP_Own_Lagged was then calculated by moving the ΔP_Own observations forward by one period and set as 0 in case the observation was the first for a stock. Appendix B includes the distribution graphs of the ΔP_Own variable and other main variables that were adjusted.

[26] These are typically cases where the fund has less than 10 000 shares of a company or less than 200 000$ invested.

Quarterly P/E, P/B, P/S, ROE and industry codes[27] were calculated and found from COMPUSTAT, by matching the CUSIP-codes[28] of each stock in sample. In the valuation measures, P was calculated from the product of common shares outstanding and the closing price of a stock in a quarter. E and S represent the trailing 12 months net income and revenue, respectively, of a company at the end of the quarter. B represents the book value of equity of a company at quarter-end.

Number of analysts was gathered from Thomson Reuters I/B/E/S -database, by flagging a company with an analyst if he/she had issued a yearly earnings per share -target for a firm. Quarterly earnings surprise was also calculated from I/B/E/S-dataset, by averaging the analysts' quarterly EPS targets and then subtracting the actual value released by the company. These were then standardized by dividing with the absolute value of the average expectation.

Daily price data for Russell 3000 -index was directly available from Yahoo Finance, which was used to calculate quarterly skewness and standard deviation -metrics for the index. Price data for individual stocks was gathered from CRSP by CUSIP-codes.

Before running the tests for H1 and H2, I purified all the other relevant variables from clear outliers and irrelevant observations: EPS surprise was set between -500% and +500% for anything above these wouldn't be very sensible, market cap was set above 50m because some stocks in the sector funds were extremely small and likely exhibit unwanted characteristics, negative deviation and IRS were capped at 20% and |5|, respectively, because a few observations were clearly out of line due to missing price data (Appendix B), PB was set below 100, PE between 0 and 1000, PS below 1000 and ROE between -200% and +200%.

The intuition behind setting limits to variables is that many of the variables would otherwise have observations that are in no way sensible and thus would affect the results in unwanted ways. Finally, I deleted quarterly observations for stock's that had had a stock split for these observations would be largely biased[29]. Additionally, it has been shown (Grinblatt et al., 1984) that stock splits tend to affect the returns of the stocks which is why I decided to exercise caution

[27] The specific industry codes used as dummies in the regressions are GICS Industry Groups (GGROUP) -codes. This was chosen for its high mapping-% with my dataset and large variability as GGROUP has 26 unique identifiers.
[28] COMPUSTAT actually uses the 9-digit CUSIP-code for stocks, so the codes from CDA/CRSP-structure had to be modified first.
[29] Split quarters were identified from CRSP using the CFACPR-variable which is a factor used to adjust the raw prices. A change in the factor within a quarter implies that there has been a split.

and remove these data points. The descriptive statistics for all the variables and their remaining observations is shown in Table 2.

Table 2: Descriptive statistics of the variables

Below are shown the descriptive statistics of the main variables used in the regressions. The data points have been combined from COMPUSTAT, CRSP, CDA, I/B/E/S, Yahoo Finance and from the AS dataset of Martijn Cremers. Some of the variables have been capped as per the discussion in the data-section.

	N	Min	Max	Average	Median	Stdev
P_Own (%)	284 402	0.00	31.11	3.38	1.18	4.31
ΔP_Own (%)	223 734	-3.99	3.99	0.16	0.03	0.93
IR (%)	19 349 554	-97	500	0.07	-0.04	15.00
IRS	281 708	-4.99	4.99	0.23	0.19	1.18
Negative deviation (%)	284 402	0.00	19.98	1.95	1.50	1.52
Stdev	284 402	0.03	4020	3.26	2.30	11.73
Mcap (m$)	239 880	50	859 968	5788	734	22 394
P/E	194 261	0.00	1000	60.00	31.90	92.53
P/B	263 614	0.00	100	3.13	1.88	5.11
P/S	239 237	0.00	1000	9.80	2.46	45.50
Volume (m)	284 402	0.13	825 293	1024	189	5314
Analysts	224 568	1	56	7	5	6.70
EPS Surprise (%)	184 730	-500	500	4	4	72
Russell Skewness	284 402	-1.50	0.82	-0.11	-0.05	0.46
Russell Stdev	284 402	0.36	4.30	1.08	0.89	0.61
Quarters	72					
Industries	26					

Identifying index deletions

Historical index compositions were gathered from COMPUSTAT and the following indices were chosen as target indices due to their prominence, firm-type variability and size: S&P 500 index, S&P 500 Growth index, S&P 500 Value index, S&P 400 Mid-Cap index, S&P 600 Small-Cap index, S&P 100 index and NASDAQ 100 index.

Russell 1000/2000 indices were not included in the sample for it is a well-known fact that once a stock is deleted from Russell 1000 it gets automatically added to Russell 2000 where its weight is a lot bigger than in Russell 1000[30]. Thus, for those cases the deletion effects would be biased. The file extracted from COMPUSTAT includes all the historical firms in the indices by CUSIP and it lists the dates on which a stock has been added or deleted from an index.

With these indices and their compositions, I identified 5050 deletions out of which 4383 instances were unique date/deletion-observations during the sample period of 2000 – 2017. The stock prices and returns were then again gathered for all of the event stocks from CRSP by the CUSIP-code of each firm.

[30] This is actually a convenient way to have exogenous variation in passive ownership as is shown in the paper by Chang et al. (2015).

V – Results

Realized idiosyncratic skewness

The results for passive ownership's effect on realized idiosyncratic skewness running Equations 3 and 4 are shown in Table 3 with the robust t-scores (White standard errors)[31] in parenthesis under each coefficient. There are three different models tested for each regression type in Table 3. These initial regressions suggest that the level of passive ownership has a statistically significant negative effect on realized skewness and that the changes and lagged changes in passive ownership further contribute to the negative effect on realized skewness once we control for a group of other variables, namely the time and industry -specific factors (Model 6). Due to the slight risk of multicollinearity in the regressions 1-2 and 4-5, indicated by the correlations (Appendix B) among the variables, I drop a group of variables in regressions 3 and 6 and consider the results of these regressions as the most precise estimates for the effects.

Thus, by my estimates (Models 3 and 6), on average the level of passive ownership decreases the quarterly realized idiosyncratic skewness by about 0.04 for each 10% out of total shares out owned by passive institutions, and that the realized skewness is further decreased by 0.01 – 0.02 for each 1% of shares outstanding purchased by the institutions in the same quarter and by a further 0.01 – 0.02 in the subsequent one. As the mean realized quarterly skewness[32] in my sample was approximately 0.23, the initial results indicate that passive ownership and the flows can in some cases push the realized skewness to the negative. The results also hold if the co-effect of the variables is not considered, as is shown in the Robustness-section of this paper (Table 10).

Somewhat surprisingly, the firm-specific factors proxying for the glamour effect do not seem to have a large impact on the realized skewness, though they are consistently significant in statistical sense. Out of the three valuation measures, the P/B ratio seems to be the most important driver of realized skewness where an increase of 1 in the ratio increases the realized skewness by around 0.005 – 0.01. The market-level skewness and standard deviation -measures are consistently significant in explaining the stock-level realized skewness but their effect

[31] I.e. by using heteroscedasticity consistent standard errors, courtesy of: https://economictheoryblog.com/2016/08/07/robust-standard-errors-in-r-function/

[32] The mean skewness is in line with previous studies such as Albuquerque (2012) who found the stock-level mean skewness to be around 0.25 – 0.30 during a similar time-period.

mostly disappears in models 4-6, implying that controlling for time captures the market-level effects quite well. The number of analysts and EPS surprise -variables are highly significant, and the EPS surprise -variable is a very prominent variable in explaining quarterly realized skewness in terms of size and statistical significance, where a 10%-point positive surprise implies a 0.02 increase in realized skewness. This finding is rather important for future studies around realized skewness, for as was mentioned earlier, there hasn't been much research around the topic and the EPS surprise variable has not been previously used in any context.

The coefficient for firm size (natural logarithm of market cap in millions), lagged skewness, lagged standard deviation and lagged returns are well in line with the findings of Chen et al. (2001) in terms of both significance and magnitude, implying that these factors are still very important nearly two decades later. Notably the R-squared values of the models (~ 2 - 4%) are similar to the models in their paper as well as the paper by Ben-David et al. (2017), affirming that R-squared values of this magnitude are typical when modelling realized stock-level skewness. Still, it of course raises a question about the goodness-of-fit for the models which is why I further plot the residuals of the main models in Appendix C. As there shouldn't be major violations of the OLS-assumptions, the coefficients shown in Table 3 are good approximations about the mean effects, regardless of the low R-squared values.

Table 3: Passive ownership and flows on the skewness of a stock's idiosyncratic returns distribution

The regressions are run for the whole time period of 2000 - 2017 which consists of a total of 284 402 individual firm-quarter findings. Realized skewness is the Fisher-Pearson skewness during each quarter calculated from idiosyncratic returns. P_Own indicates the quarterly ownership-% of all passive funds out of total shares out for each company. ΔP_Own indicates the quarterly percentage point change in the ownership of passive funds for each stock in the sample. ΔP_Own_Lagged is the change during the previous quarter. Russ_Skewness and Stdev indicate the realized returns distribution descriptors for the Russell 3000 -index during each quarter. P/E, P/B and P/S variables indicate firm-specific valuation metrics, calculated each quarter by dividing a firm's market cap with its realized 1-year net result/book value of equity/revenue. Return is the stock's return in the previous quarter. Market cap is the quarter end market cap of a stock measured in millions. Volume is the amount of sells of a stock in a quarter, measured in thousands. Analysts are measured as the number of analysts giving yearly EPS estimates for a firm. EPS surprise is the relative surprise of the EPS released in the quarter being analyzed. Lagged variables are calculated from the observations of the previous quarter. Time and sector dummies indicate the amount of dummy variables in the model. The number of observations is not stable in each model for there is lacking data for some variables, most notably the valuation metrics. The t-scores are calculated using robust standard errors and they are represented in brakets below the coefficients.

Dependent variable: Realized Quarterly Idiosyncratic Skewness	Standard OLS regression (eq 5)			Fixed-effects models (eq 6)		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.28	0.53	0.30	0.34	0.37	0.19
	(83.71)	(12.21)	(20.10)	(15.49)	(5.18)	(6.57)
P_Own	-0.012	-0.001	-0.004	-0.015	0.005	-0.004
	(-17.33)	(-0.60)	(-3.78)	(-15.25)	(1.51)	(-3.48)
ΔP_Own	-0.01	-0.01	-0.013	-0.022	-0.014	-0.022
	(-1.89)	(-2.04)	(-3.38)	(-5.83)	(-1.82)	(-4.53)
ΔP_Own_Lagged	-0.01	0.001	-0.010	-0.018	-0.004	-0.019
	(-1.59)	(0.64)	(-2.39)	(-4.98)	(-0.66)	(-3.89)
Russ_Skewness		-0.11	-0.10		0.04	-0.05
		(-10.53)	(-11.19)		(1.01)	(-0.90)
Russ_Stdev		-0.04			0.009	
		(-5.54)			(0.23)	
P/E		0.0002			0.0001	
		(4.53)			(2.70)	
P/B		0.005	0.01		0.004	0.006
		(4.53)	(12.08)		(4.54)	(9.75)
P/S		0.002			0.002	
		(4.74)			(4.46)	
ln(Market cap (m))		0.06	-0.03		0.09	-0.022
		(12.40)	(-13.10)		(18.00)	(-10.60)
ln(Volume (t))		-0.07			-0.10	
		(-15.30)			(-19.26)	
Return_Lagged		-0.003			-0.003	
		(-15.03)			(-15.90)	
Skewness_Lagged		0.014	-0.002		0.012	-0.006
		(4.19)	(-0.70)		(3.42)	(-2.30)
Stdev_Lagged		0.04	0.01		0.052	0.008
		(17.31)	(10.04)		(18.39)	(8.79)
Analysts		-0.002			-0.004	
		(-2.35)			(-4.40)	
EPS Surprise		0.29	0.18		0.29	0.18
		(37.30)	(35.90)		(36.93)	(35.79)
Number of time dummies**				71	71	71
Number of sector dummies**				25	25	25
N	204 828	80 882	118 836	204 828	80 882	118 836
Adj. R^2	0.18%	2.6%	1.5%	0.13%	3.7%	2.2%

** Most sector and time dummies are highly significant in all models.

Realized idiosyncratic negative deviation

Hypothesis 1 is further supported by the regressions results in Table 4, where I explain realized negative idiosyncratic deviation, i.e. the negative deviation of the firm-specific returns, with the passivity variables and a group of control variables (Equation 7). I found that both the change and the lagged change in passive ownership have a consistent positive effect on the idiosyncratic negative deviation, where a 1%-point quarterly increase in passive ownership leads to an increase of 0.01 – 0.04%-points in realized negative deviation in the same quarter and an additional 0.01 – 0.04%-points in the subsequent quarter (Models 3 to 5). The statistical significance for the same quarter -effects varies from model to model and does not reach a high significance in all instances whereas the lagged change is highly significant across the models. The result is intuitive as I would expect that it takes some time to correct the effects of passivity.

Interestingly though, the *level* of passive ownership has an extremely significant negative effect on realized negative deviation (1%-points increase in level predicts a 0.03%-point decrease in negative deviation), implying that stocks that have the highest level of ownership by passive money tend to have less prominent left tail in the returns distribution. This result is actually intuitive and in line with prior findings where it has been shown that the most heavily indexed stocks co-move more together and behave more like the index itself, thus having less extreme negative or positive idiosyncratic movements overall.

The result that passive ownership has a negative coefficient for both realized skewness and negative deviation do not contradict each other, for it simply means that the level of passive ownership tends to lead to less prominent tail movements overall. Similarly, the results for the flows, where the coefficients were the opposite, implies that the flows lead to higher negative movements, i.e. negative skewness due to more prominent left-tail movements, hence supportive of H1.

The effects of passive flows and level of ownership are a lot smaller than many of the other variables such as index-level movements, EPS surprise or lagged stock-specific movements. Also, the results for the control variables are intuitive, i.e. smaller negative deviation predicted by size and EPS surprise, and larger by lagged risk measures (standard deviation and skewness). Regardless of the magnitude of the passivity variables, the fact that the flows consistently predict an elevated negative deviation is an important finding regarding the changes to the pricing process caused by passive institutions in the short-term and it ultimately implies that active investors are correcting the potential mispricing caused by passive flows, at least to some extent.

The dependent variable in the regressions 1-5 is the negative deviation calculated from firm-specific returns. This may pose a problem in the sense that during a quarter the amount of negative IR days for individual stocks may be quite low. Therefore, in Model 6, I switch the dependent variable to negative deviation calculated from the total returns, i.e. returns from which the market return has not been detracted from. The results from this regression are mostly in line with the results in Models 1-5, but there is a significant change in the coefficient for same quarter change in passive ownership.

More specifically, in Model 6, a 1%-point increase in passivity predicts a 0.017%-point *decrease* in realized negative deviation. This may be due to the tendency for stocks to trend upwards and the fact that passive flows have been mostly positive during the sample period, although market and time -specific factors have been controlled for. The result is also consistent with the notion that passive flows can have positive effects on stock returns initially which can lead to decreased left-tail moves in terms of total returns, and that corrections take time (positive lagged coefficient).

Perhaps the most controversial interpretation of these results can be derived from the fact that same quarter flows predict smaller negative deviation from non-adjusted returns but larger negative deviation for the firm-specific component of returns. Therefore, one could conclude that the market component of returns is slower in correcting the effects caused by passive flows. The result implies that the effects of passivity may not have as meaningful effects on the relative valuations among companies (firm-specific returns adjust more quickly), but more on the aggregate stock market valuation. This would also be consistent with the fact that there are now more indices than stocks, i.e. most stocks are indexed, and that passivity takes no view on corporate valuations, so the effects of passivity could perhaps be more severe at the market level, instead of at the company-level. Nevertheless, the more likely driver behind the results is the fact that stocks tend to trend upwards and the passive flows have coincided with this, but it is an interesting finding in any case.

Ultimately these results give some explanation for the disappearance of stock return randomness found in the papers by Qin and Singal (2015) and Coles et al. (2018) and it is consistent with the findings of Baltussen et al. (2017) regarding the index-level negativity in autocorrelation in recent years.

Table 4: Passive ownership and flows on the idiosyncratic negative deviation of a stock's returns distribution

The regressions are run for the whole time period of 2000 - 2017 which consists of a total of 284 402 individual firm-quarter findings. The dependent variable is the realized negative deviation, calculated from idiosyncratic returns. P_Own indicates the quarterly ownership-% of all passive funds out of total shares out for each company. ΔP_Own indicates the quarterly percentage point change in the ownership of passive funds for each stock in the sample. ΔP_Own_Lagged is the change during the previous quarter. Russ_Skewness and Stdev indicate the realized returns distribution descriptors for the Russell 3000 -index during each quarter. P/E, P/B and P/S variables indicate firm-specific valuation metrics, calculated each quarter by dividing a firm's market cap with its realized 1-year net result/book value of equity/revenue. Return is the stock's return in the previous quarter. Market cap is the quarter end market cap of a stock measured in millions. Volume is the amount of sells of a stock in a quarter, measured in thousands. Analysts are measured as the number of analysts giving yearly EPS estimates for a firm. EPS surprise is the relative surprise of the EPS released in the quarter being analyzed. Lagged variables are calculated from the observations of the previous quarter. Time and sector dummies indicate the amount of dummy variables in the model. The number of observations is not stable in each model for there is lacking data for some variables, most notably the valuation metrics. The t-scores are calculated using robust standard errors.

Dependent variable: Realized Quarterly Idiosyncratic Negative Deviation						**Pure returns**
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	2.16	0.39	0.32	0.33	1.26	0.75
	(460)	(14.71)	(12.85)	(14.95)	(35.04)	(18.26)
P_Own	-0.07	-0.03	-0.03	-0.03	-0.03	-0.03
	(-88.01)	(-43.98)	(-23.73)	(-29.15)	(-30.05)	(-27.38)
ΔP_Own	0.051	0.044	0.025	0.042	0.007	-0.017
	(0.91)	(3.75)	(1.53)	(3.41)	(0.92)	(-5.18)
ΔP_Own_Lagged	0.028	0.024	0.025	0.035	0.013	0.014
	(1.30)	(4.59)	(5.87)	(10.33)	(4.17)	(4.95)
Russ_Skewness		0.14	0.13	0.33	0.26	0.32
		(24.81)	(19.24)	(34.72)	(36.73)	(44.60)
Russ_Stdev		0.44	0.37			
		(76.62)	(46.22)			
P/E		0.0003	0.0002			
		(6.43)	(4.52)			
P/B		0.010	0.009	0.013	0.003	-0.00
		(11.65)	(8.17)	(8.06)	(1.30)	(-0.76)
P/S		0.005	0.003			
		(4.13)	(3.32)			
ln(Market cap (m))		-0.41	-0.34	-0.41	-0.40	-0.49
		(-124)	(-35.62)	(-38.48)	(-56.89)	(-49.46)
ln(Volume (t))		0.29	0.24			
		(83.00)	(30.21)			
Return_Lagged			-0.002			
			(-2.65)			
Skewness_Lagged			0.031	0.017	0.011	0.016
			(6.79)	(6.96)	(8.77)	(8.07)
Stdev_Lagged			0.11	0.030	0.034	0.024
			(5.76)	(1.82)	(1.74)	(1.75)
Analysts		0.002	0.001			
		(1.12)	(1.05)			
EPS Surprise		-0.06	-0.061	-0.10	-0.10	-0.09
		(-11.80)	(-12.67)	(-23.17)	(-23.02)	(-20.80)
Number of time dummies**					71	71
Number of sector dummies**					25	25
N	207 828	82 927	81 882	119 836	119 836	119 836
Adj. R^2	3.8%	39.1%	42.8%	23.2%	43.6%	48.4%

** Most sector and time dummies are highly significant

Cross-sectional differences of realized skewness and negative deviation

Realized skewness regression results from grouping the sample in to five groups based on their most recent released ROE-value, which is used as a proxy for the overall economic fundamentals of a company, are shown in Table 5, where Group 1 indicates the group with the worst fundamentals and Group 5 the one with the best fundamentals.

Passivity's effect on realized skewness for different groups seems to vary a bit, namely that the effects are most prominent for the group of stocks with the lowest profitability both in terms of the level and flows of passivity. Comparing the two extremes, the level coefficient has a 10 times more negative effect on realized skewness for the group with the worst fundamentals (-0.02 vs -0.002). The flow variables are notably different as well, with -0.05 vs -0.02 in the same quarter and -0.06 vs +0.02 in the subsequent quarter, for worst ROE group and best ROE group, respectively. The effects of passivity are also notably more statistically significant (absolute t-values > 4.00) for the group with the worst fundamentals compared to the other groups.

These results are consistent with the notion that stocks that have worse fundamentals wouldn't perhaps otherwise be in much demand and hence the effects caused by passive flows may be elevated, whereas the stocks with the best fundamentals would receive the buying pressure in any case and thus the reversal effects are less prominent. Or in other words, the share of non-fundamental demand is perhaps higher for stocks that perform poorly economically and thus leads to larger effects on the downside once the effects are corrected.

Therefore, the results from H2 not only reveal that cross-sectional differences are significant, but also support the notion that passive share matters because it is intuitive to think that the share of non-fundamental trading is higher for the economically worst companies.

The control variables are mostly stable across groups and the coefficients are in line with the results of the total sample in Table 3.

Table 5: Passivity on realized idiosyncratic skewness, grouped by ROE

The regressions are run for five different groups over 2000 - 2017 based on the latest revealed return of equity -value (ROE) of the firm. The dependent variable is the realized skewness, calculated from idiosyncratic returns. P_Own indicates the quarterly ownership-% of all passive funds out of total shares out for each company. ΔP_Own indicates the quarterly percentage point change in the ownership of passive funds for each stock in the sample. ΔP_Own_Lagged is the change during the previous quarter. Russ_Skewness and Stdev indicate the realized returns distribution descriptors for the Russell 3000 -index during each quarter. P/B variable indicates firm-specific valuation metric, calculated each quarter by dividing a firm's market cap with its realized book value of equity. Return is the stock's return in the previous quarter. Market cap is the natural logarithm of the quarter end market cap of a stock measured in millions. EPS surprise is the relative surprise of the EPS released in the quarter being analyzed. Lagged variables are calculated from the observations of the previous quarter. Time and sector dummies indicate the amount of dummy variables in the model. The number of observations is not stable in each model for there is lacking data for some variables, most notably the valuation metrics. The t-scores are calculated using robust standard errors.

Dependent variable: Realized Quarterly Idiosyncratic Skewness	Worst ROE				Best ROE
	Group 1	Group 2	Group 3	Group 4	Group 5
Intercept	0.54	-0.04	-0.13	-0.23	-0.04
	(6.60)	(-0.49)	(-2.03)	(-4.30)	(-0.61)
P_Own	-0.02	-0.01	0.002	0.004	-0.002
	(-4.22)	(-3.32)	(0.87)	(1.41)	(-0.73)
ΔP_Own	-0.05	-0.02	-0.01	-0.01	-0.02
	(-4.09)	(-1.00)	(-0.60)	(-0.99)	(-2.00)
ΔP_Own_Lagged	-0.06	-0.02	-0.02	-0.01	0.02
	(-4.60)	(-1.44)	(-2.05)	(-0.94)	(1.89)
Russ_Skewness	-0.13	0.15	-0.09	-0.12	0.18
	(-0.89)	(0.94)	(-0.67)	(-1.08)	(1.60)
P/B	0.007	0.03	0.03	0.02	0.004
	(5.53)	(5.90)	(6.54)	(5.32)	(3.30)
ln(Market cap (m))	-0.02	-0.01	0.002	-0.004	-0.01
	(-3.53)	(-0.97)	(0.53)	(-0.93)	(-2.21)
Return_Lagged	-0.002	-0.003	-0.003	-0.003	-0.003
	(-8.50)	(-8.52)	(-8.15)	(-8.70)	(-7.75)
Skewness_Lagged	0.03	0.02	-0.002	0.01	0.009
	(4.30)	(2.70)	(-0.30)	(2.60)	(1.70)
Stdev_Lagged	0.01	0.03	0.03	0.03	0.04
	(7.40)	(7.88)	(7.45)	(6.80)	(6.70)
EPS Surprise	0.09	0.15	0.27	0.39	0.32
	(10.00)	(17.33)	(20.60)	(24.00)	(18.76)
Number of time dummies**	71	71	71	71	71
Number of sector dummies**	25	25	25	25	25
N	18 403	20 531	25 925	27 857	26 885
Adj. R^2	2.9%	2.9%	3.2%	3.8%	2.8%

** Most sector and time dummies are highly significant

Similarly, as in the H1 tests, to study the results further and understand the nature of the effects on realized skewness, I separately run negative deviation -regressions for each group in Table 6.

The results give support to the discussion above, because the group with the worst fundamentals has the greatest coefficients for both the flow variables (0.015 vs 0.009 and 0.041 vs 0.022), implying that for these stocks the non-fundamental buying pressure leads to the most extreme left-tail movements.

The level of passivity again has a consistently negative effect on realized negative deviation as was discussed in the H1 results earlier. However, the effect seems to be slightly more prominent for the stocks with the best fundamentals. The story here may be that "good" stocks behave more like their respective indices (have less extreme tail movements) than the "bad" stocks because they are more prominent in the indices and tend to have higher passive levels overall. The difference of the magnitude for the coefficient among the extremes is not very large though (-0.01 vs -0.03).

Ultimately, these regressions highlight the importance of firm characteristics when their sensitivity to passive non-fundamental pressure is analysed. The ROE variable seems to be a rather meaningful variable in grouping stocks, but there are surely endless amounts of additional variables as well. For future research, one can further study these sorting variables.

Table 6: Passivity on realized negative deviation, grouped by ROE

The regressions are run for five different groups over 2000 - 2017 based on the latest revealed return of equity -value (ROE) of the firm. The dependent variable is the realized negative deviation, calculated from idiosyncratic returns. P_Own indicates the quarterly ownership-% of all passive funds out of total shares out for each company. ΔP_Own indicates the quarterly percentage point change in the ownership of passive funds for each stock in the sample. ΔP_Own_Lagged is the change during the previous quarter. Russ_Skewness and Stdev indicate the realized returns distribution descriptors for the Russell 3000 -index during each quarter. P/B variable indicates firm-specific valuation metric, calculated each quarter by dividing a firm's market cap with its realized book value of equity. Return is the stock's return in the previous quarter. Market cap is the natural logarithm of the quarter end market cap of a stock measured in millions. EPS surprise is the relative surprise of the EPS released in the quarter being analyzed. Lagged variables are calculated from the observations of the previous quarter. Time and sector dummies indicate the amount of dummy variables in the model. The number of observations is not stable in each model for there is lacking data for some variables, most notably the valuation metrics. The t-scores are calculated using robust standard errors.

Dependent variable: Realized Quarterly Idiosyncratic Negative Deviation	Worst ROE				Best ROE
	Group 1	Group 2	Group 3	Group 4	Group 5
Intercept	4.99	2.82	2.83	2.77	2.69
	(54.00)	(18.07)	(26.90)	(22.61)	(17.43)
P_Own	-0.01	-0.01	-0.02	-0.02	-0.03
	(-2.66)	(-6.78)	(-11.10)	(-11.76)	(-13.02)
ΔP_Own	0.015	-0.015	-0.016	-0.008	0.009
	(1.34)	(-2.49)	(-2.91)	(-1.38)	(1.33)
ΔP_Own_Lagged	0.041	0.009	0.010	0.025	0.022
	(4.34)	(1.64)	(2.05)	(7.04)	(5.12)
Russ_Skewness	0.32	0.18	0.19	0.18	0.25
	(13.20)	(11.42)	(14.07)	(12.94)	(16.20)
P/B	-0.006	-0.001	0.011	0.012	-0.002
	(-5.50)	(-0.20)	(2.50)	(3.20)	(-3.51)
ln(Market cap(m))	-0.28	-0.14	-0.15	-0.15	-0.14
	(-39.40)	(-20.92)	(-29.20)	(-23.47)	(-17.40)
Return_Lagged	-0.003	-0.005	-0.005	-0.003	0.00
	(-8.88)	(-7.75)	(-5.50)	(-2.49)	(0.24)
Skewness_Lagged	0.05	0.04	0.04	0.03	0.03
	(5.63)	(6.47)	(5.84)	(5.19)	(7.01)
Stdev_Lagged	0.02	0.14	0.09	0.11	0.17
	(3.04)	(5.70)	(4.62)	(4.46)	(6.09)
EPS Surprise	-0.09	-0.06	-0.08	-0.06	-0.05
	(-8.90)	(-9.79)	(-7.39)	(-4.17)	(-3.49)
Number of time dummies**	71	71	71	71	71
Number of sector dummies**	25	25	25	25	25
N	18 403	20 531	25 925	27 857	26 885
Adj. R^2	29.9%	40.6%	36.8%	35.4%	38.9%

** Most sector and time dummies are highly significant

Index deletion event window formation

Before testing the third hypothesis of the thesis, i.e. whether index-deletion effects are a function of the level of passive ownership, I first merged the passive ownership -dataset described earlier with the event study -dataset and deleted events for which I don't have ownership data. A total of 1982 events were identified for these tests. For initial tests, I then formed 5 groups based on the level of passive ownership in the previous quarter to test whether the amount of passive ownership prior to an index deletion is associated with the abnormal returns of stocks around the deletion events, where Group 1 indicates the stocks with the least passive ownership. Summary statistics of the five groups are shown in Table 7 indicating that there are no significant differences in a few key variables.

Table 7: Summary statistics for event study groups

	Passivity Groups	Mean	Min	Max	Stdev	Mean time
Market Cap						
Least passive	**1**	20 047	18	438 702	35 676	08/2011
	2	13 663	8	364 064	32 367	09/2004
	3	15 263	13	439 013	43 458	07/2006
	4	18 497	11	276 808	34 272	11/2010
Most passive	**5**	10 003	14	140 018	14 005	07/2015
P/B						
	1	4,82	0,08	245,70	13,38	
	2	4,19	0,08	141,50	9,19	
	3	4,53	0,04	359,08	18,66	
	4	4,61	0,02	491,00	26,41	
	5	3,97	0,06	172,80	9,40	
Analysts						
	1	15,45	1	45	8,17	
	2	12,45	1	37	8,03	
	3	11,11	1	38	6,98	
	4	14,28	1	46	8,42	
	5	12,14	1	37	8,17	
EPS Surprise						
	1	0,03	-3,32	4,90	0,39	
	2	-0,06	-6,49	1,83	0,79	
	3	-0,05	-5,37	1,96	0,65	
	4	0,02	-12,39	2,76	0,73	
	5	0,05	-5,89	7,62	0,71	
ROE						
	1	-0,03	-34,33	14,32	2,50	
	2	0,01	-12,83	5,88	0,98	
	3	1,66	-13,79	16,91	34,71	
	4	1,14	-18,21	33,74	23,46	
	5	-0,03	-38,72	6,82	1,84	

Additionally, the index deletion -events allowed me to further check if the passivity sample and its change has been constructed properly, for I know for a fact that index-deletions lead to decreased passive ownership, ceteris paribus.

Therefore, in Figure 4 A I plot the mean change of passive ownership and in B the mean change standardized by the level of passive ownership from 4 quarters prior to the deletion quarter to 4 quarters after the deletion quarter. These figures clearly indicate that the passive ownership -variable behaves exactly as it should in the case of index deletions and therefore verifies the dataset overall, i.e. there is a significantly large decrease in the share owned by passive institutions in the quarter where the index deletion came in to effect.

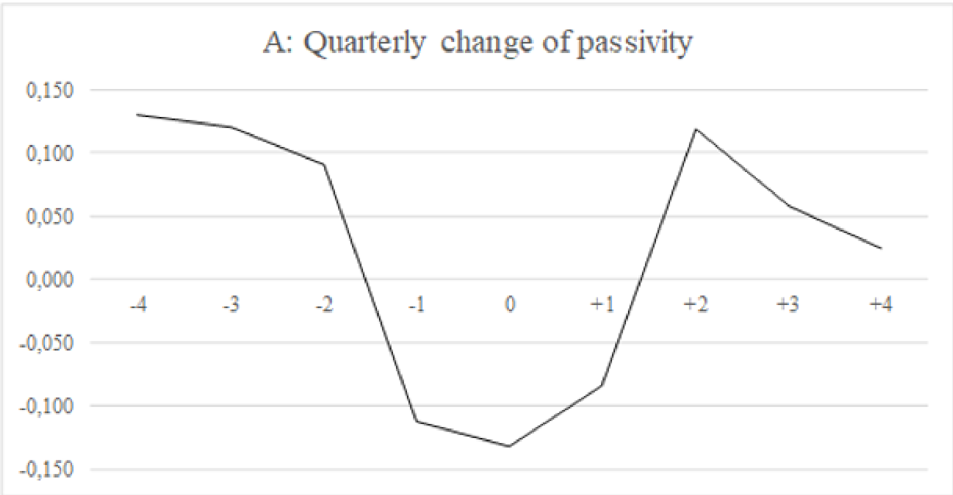


Figure 4 A: Change of passive ownership in index deletions

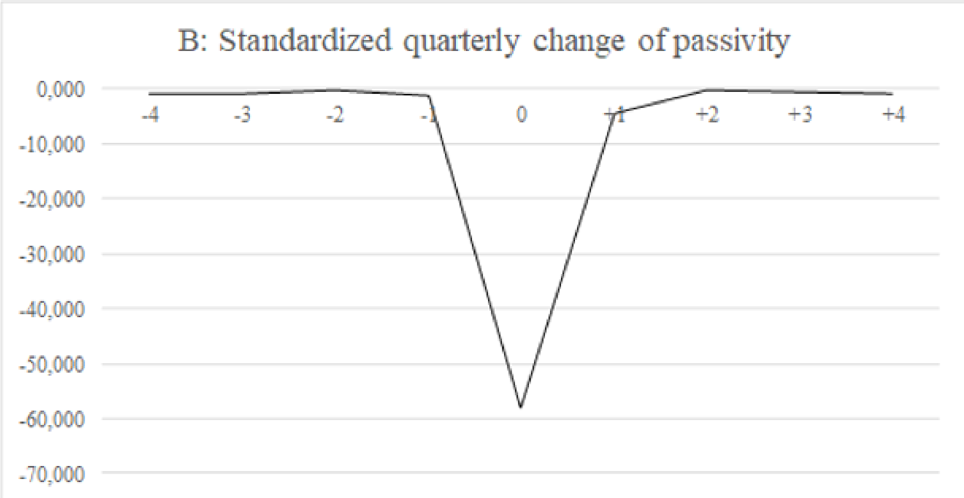


Figure 4 B: Change of passive ownership in index deletions

The total event study -period under main analysis is from 50 days prior to the index deletion to 50 after the deletion. The long period is chosen because I decided to form four distinct sub-periods in the event studies and compare their mean abnormal returns: 1st from -50 to -35 days which indicates the period in which the event was not yet known, implying that any trading done here was only by the most active and highest risk-tolerating investors anticipating the announcement, 2nd period from -34 to -5 day proxies for the announcement period[33], where also traders with lower risk preference and probably less sophistication are willing to trade the deletion, 3rd from -4 to +5 which is the period in which index funds are expected to make their adjustments and it proxies for the non-fundamental trading period and finally 4th from +6 to +50 which indicates the post-deletion period. The period -50 - +5 is also calculated for it ignores the post-deletion rebound that is apparent in the results.

Index deletion event study results

Table 8 panel A shows the results for the cumulative abnormal returns separated by period/level of passivity and the results are also plotted in Figure 5 with daily CAAR and its 95% confidence interval[34]. The effects caused by index deletion are clearly extremely large and tend to be negative. Importantly, the group with the least passive ownership is strikingly different

[33] For example, the S&P500 reports index deletions 1-30 days prior to the actual effective date.
[34] The event studies were run using R package "Eventstudies" created by Ajay Shah, Chirag Anand, Vikram Bahure and Vimal Balasubramaniam.

compared to the other groups in all of the periods under analysis and the results suggest that for this group the event effects are actually slightly positive. This finding may be explained with the notion that if a stock has a very low level of passive ownership prior to a deletion, its weight in the index is likely also very small and any trading related to it being dropped may have already been conducted earlier. Whereas stocks that still have moderate to high amounts of passive ownership have more mixed views about their future in the index and thus the deletion is a surprise to at least some market participants.

Importantly, the clear distinction of the results between groups 1 and 5 is rather strong evidence that the sample building process for the passive ownership -variables is successful and it brings robustness to the results of the earlier regressions.

The most interesting finding in the deletions is the trend in the period -50 to -35 when the event windows are run by passivity group, which indicates that the more passive ownership prior to the deletion, the higher the negative abnormal returns are in the anticipation window. The anticipation window represents the period where the index deletion was still unknown to the markets which implies that higher abnormal negative returns in this period suggests that anticipators view these stocks as being less accurately priced and are willing to trade them more heavily. And if it is assumed that there are indeed investors that possess better skills in picking stocks and anticipating future index compositions, then I would expect these investors to be the ones trading before anyone else. If their views are indeed accurate, it would suggest that the level of passive ownership and new flows may be driving the valuations, not only the mere fact that a stock is in an index.

These anticipation effects are discussed also in a related recent study by Arnott et al. (2018) where they show that for stocks in the S&P 500 since 1989, the one-month *prior* return over the market return has been +1.84% and -6.57% for additions and deletions, respectively. They discuss that the most plausible explanation is that there is substantial amount of smart money anticipating index composition changes and trading the stocks accordingly. The results for the deletion event studies in this thesis highly support this notion.

Other than the observations of the first group and in the anticipation period, there doesn't seem to be a clear trend where higher passivity would lead to larger effects in the other windows. From -50 to +50 the group with the most passive ownership has the largest negative CAAR and the only group with a consistently negative 95% confidence interval throughout the event window, but the results are not very consistent for the group with 2nd most passive ownership is only slightly negative. It would actually seem that the effects in the windows from -50 to +6 are the biggest for stocks with around the mean amount of passive ownership.

Perhaps the effects of the deletion on stock prices from the announcement onwards have little to do with the potential effects that passive money has had and more with the mere fact that there is a deletion and speculators trade the stocks in a similar fashion regardless of the share owned by passive institutions, hence leading to larger than warranted price decreases that are then corrected to some extent, as seen by the apparent rebound in the CARs in Figure 5.

Nevertheless, the finding that the least passive stocks deviate so much from the rest of the sample and that the anticipation effects have a clear trend with respect to passivity, indicate that there may be latent long-term premiums caused by passive money.

Also, to my knowledge this type of event study by group and by window has not been studied in this context yet, and the result that there are clearly a lot of variation with respect to index deletions is an interesting finding and offers various research topics for the future. One could for example analyse the short interest for firms prior to the deletion announcement to see whether smart money is indeed trading the announcements and if it is linked to the amount of passive ownership.

Table 8: Abnormal returns around index deletions grouped by passivity and year

This table shows the results for the event studies run around index deletions for S&P 500 index, S&P 500 Growth index, S&P 500 Value index, S&P 400 Mid-Cap index, S&P 600 Small-Cap index, S&P 100 index and NASDAQ 100 index from 2000 to 2017. Abnormal returns are calculated as the excess returns over the Russell 3000 -index. In Panel A, five studies are run by grouping 1982 deletions by the level of passive ownership in the previous quarter. In Panel B, the events are grouped by the year in which the deletion became effective. For Panel A results, the significance intervals can be seen in Figure 2 where the values indicate the CAAR at each time point and its 95% confidence interval. Panel B significance results at the 95% level are indicated with *.

Panel A: Grouped by passivity

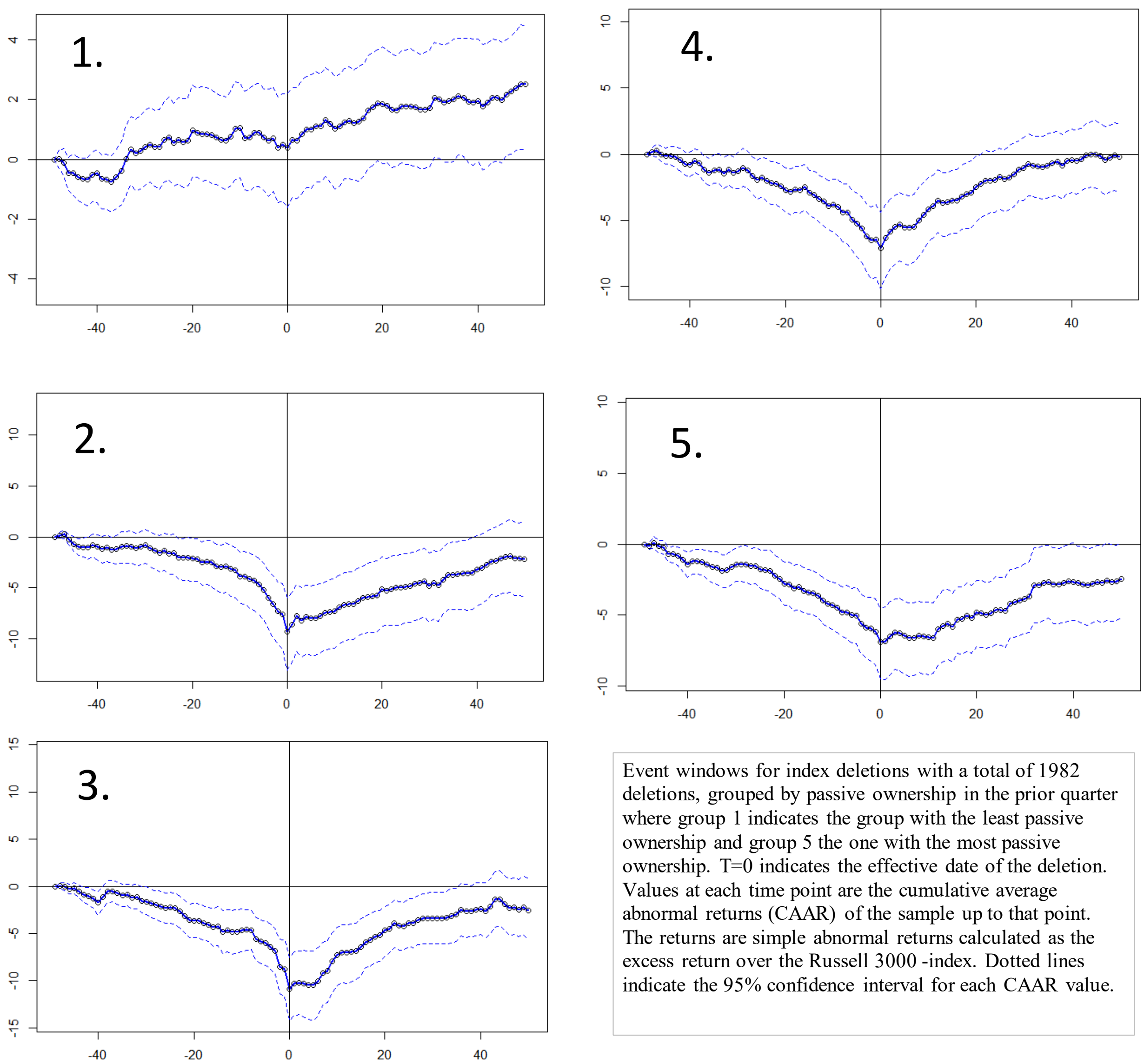
		Mean abnormal return per period							
		-50 / -35	-34 / -5	-4 / +5	+6 / +50	-50 / +50	-50 / +5	Deletion date AR	Mean lagged P_Own
Least passive	**1**	0,20	0,46	0,34	1,42	2,75	1,23	0,02	0.12%
	2	-0,72	-4,07	-1,83	5,60	-1,83	-7,46	-1,31	1.33%
	3	-0,75	-4,31	-2,68	5,70	-2,07	-8,16	-1,46	4.00%
	4	-1,14	-4,08	-0,23	5,21	-0,41	-5,66	-0,47	7.63%
Most passive	**5**	-1,63	-3,74	-0,99	4,45	-2,14	-6,48	-0,63	13.94%

Panel B: Grouped by year

	Mean abnormal return per period						
	-50 / -35	-34 / -5	-4 / +5	+6 / +50	-50 / +50	-50 / +5	Deletion date AR
2000	1,81	-6,37	-3,47	5,94	-1,53	-8,12	-2,83
2001	-1,17	-11,43	-3,54	2,71	-12,79	-17,34	-22,43
2002	-5,22	-7,57	-7,17	9,97	-11,80	-21,39	-20,64
2003	1,87	-2,37	1,05	6,96	7,16	0,02	-1,55
2004	-0,49	-3,39	0,85	3,59	-0,87	-4,12	-4,75
2005	-1,62	-2,38	-0,48	-1,09	-5,20	-4,37	-5,84
2006	-5,94	-6,32	-0,04	-1,76	-18,16	-15,11	-16,13
2007	1,22	1,35	1,48	-0,45	3,82	3,84	3,81
2008	-0,77	-12,03	-8,13	14,77	-7,43	-21,54	-21,30
2009	-2,39	-8,80	0,71	21,82	14,82	-7,42	-13,14
2010	-1,92	2,91	0,59	5,17	6,60	1,35	1,80
2011	1,08	-0,49	0,97	-5,89	-4,07	1,17	-1,13
2012	-3,38	-1,25	-0,33	6,13	-0,14	-6,85	-7,38
2013	-3,14	-3,20	0,73	-3,03	-8,77	-5,67	-7,19
2014	-2,72	-4,02	-0,29	0,93	-6,72	-8,02	-7,77
2015	-0,78	-5,97	-1,25	-6,89	-15,06	-8,69	-8,14
2016	-0,05	0,37	0,46	2,04	3,24	1,33	1,17
2017	-2,95	-5,79	0,25	-0,27	-9,72	-9,71	-11,39

Additionally, I grouped the individual events by year and ran similar event studies in Panel B. The yearly events were tested because the exact level of passive ownership may not be accurate in some cases due to missing data, as was discussed in the data section of the sample building.

The yearly separation has a favourable property in the sense that I know for a fact that the mean passive ownership among all firms in year t+1 is higher than in t+0, as can be seen from Table 1. The trade-off here is that there are likely time-specific factors driving the results during an event window, i.e. events during year t+1 can systematically be different from t+0 due to various other reasons than just the level of passive ownership. Similarly though, the anticipation window from -50 to -35 has been systematically negative and somewhat more consistent in the latter years which is further evidence that the level of passivity may be driving valuations that are then corrected by active investors once they anticipate an index deletion event in the near future. But overall, the results in Panel B do not indicate that there would be a clear pattern in the evolution of deletion effects over the years.



Event windows for index deletions with a total of 1982 deletions, grouped by passive ownership in the prior quarter where group 1 indicates the group with the least passive ownership and group 5 the one with the most passive ownership. T=0 indicates the effective date of the deletion. Values at each time point are the cumulative average abnormal returns (CAAR) of the sample up to that point. The returns are simple abnormal returns calculated as the excess return over the Russell 3000 -index. Dotted lines indicate the 95% confidence interval for each CAAR value.

Figure 5: Index-deletion CAAR windows by passive ownership groups

Regression results for abnormal returns in index-deletions

Further analysis about passive ownership's effects on abnormal returns around deletions is included in Table 9 where various regressions have been run. The dependent variable in these models is either the anticipation window CAR (-50 to -35 days) or the total window CAR (-50 to +50 days) at the stock level. Independent variables include some of the same variables as in H1 and H2 regressions, as well as the one-year market return (Russell 3000) prior to the deletion in question, the pure one-year return of the stock itself and six index dummies. The return variables have been adjusted so that they do not include any of the observations from the CAR periods in question.

The results from these regressions support the evidence from Table 8 and Figure 5, i.e. that the effect of passivity on the realized CAR in the anticipation window is consistently negative across models and the effect is at least semi-strong in statistical sense with a t-value of -1.81 in Model 3, where a 1%-point added passive ownership predicts a higher negative anticipation CAR of -0.24%. This result gives further support to the possibility of more prominent smart money trading against stocks that have had a higher passive share, consistent with the notion that passive flows can cause prices to deviate from their fundamental values over longer time periods. Due to missing data in many of the key variables, the sample size in the regressions varies somewhat though.

The mean results of passivity's effect in the total window CAR are highly negative as well (-0.29% to -0.40%) although they do not reach a very high statistical significance (t-values -0.99 to -1.24).

The regressions were also tested with year dummies outside of the results shown in Table 9 and the magnitude of the mean effects from the lagged level of passivity were unchanged, although the statistical significance decreased with new t-values being -1.2 and -1.0 for anticipation window and total window, respectively. The results of the regressions are thus not definitive but at least moderately support the third hypothesis of the thesis.

The control variables improve the models and the most noteworthy ones are the previous 1-year returns of the market and the stock itself and the size of the company.

Table 9: Passive ownership and the cumulative abnormal returns in index deletions

The regressions are run for the whole time period of 2000 - 2017 which consists of a total of 1982 individual index deletions with data for the regressions. The indices included are: S&P 500 index, S&P 500 Growth index, S&P 500 Value index, S&P 400 Mid-Cap index, S&P 600 Small-Cap index, S&P 100 index and NASDAQ 100 index. The dependent variable is the firm-level cumulative abnormal return (return over the Russell 3000) over the specified period. P_Own_Lagged indicates the quarterly ownership-% of all passive funds out of total shares out for each company in the quarter prior to the deletion. P/B and P/S variables indicate firm-specific valuation metrics, calculated each quarter by dividing a firm's market cap with its realized book value of equity/revenue. Market cap is the natural logarithm of the quarter end market cap of a stock measured in millions. Volume is the natural logarithm of the amount of sells of a stock in a quarter, measured in thousands. Analysts are measured as the number of analysts giving yearly EPS estimates for a firm. Index dummies indicate the dummy variables for the 7 indices under analysis. Russell 3000 and stock returns are the 1-year simple returns prior to the index deletion effective date. The number of observations is not stable in each model for there is lacking data for some variables, most notably the valuation metrics. The t-scores are calculated using robust standard errors.

Dependent variables:	**Anticipation window: CAR -50 / -35**			**Total window: CAR -50 / +50**		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-2.57	0.26	-11.96	-6.05	-12.98	-18.3
	(-1.90)	(0.03)	(-0.98)	(-1.88)	(-0.51)	(-0.62)
P_Own_Lagged	-0.003	-0.17	-0.24	0.07	-0.29	-0.40
	(-0.06)	(-1.35)	(-1.81)	(0.55)	(-0.99)	(-1.24)
Russ_1y_Return			-0.02			-0.35
			(-0.51)			(-3.44)
Stock_1y_Return		0.04	0.03		0.05	0.09
		(2.13)	(1.79)		(1.12)	(2.08)
P/B			0.10			-0.05
			(0.64)			(-0.13)
P/S			0.01			-0.00
			(0.10)			(-0.00)
ln(Market cap (m))		2.45	2.44		7.78	7.00
		(3.83)	(3.32)		(4.84)	(3.93)
ln(Volume (t))		-1.32	-0.36		-3.60	-2.16
		(-1.81)	(-0.42)		(-1.96)	(-1.02)
Analysts			-0.09			-0.13
			(-0.70)			(-0.39)
Index dummies	6	6	6	6	6	6
N	1 982	565	491	1 982	565	491
Adj. R^2	0.1%	10.7%	9.2%	2.3%	12.9%	12.1%

If it truly is the case that the level of passive ownership is the factor behind the results found in Table 8 and Table 9 due to added passivity's effect on corporate valuations, there are interesting implications about the most heavily indexed stocks still included in the indices. As the deleted stocks typically have smaller weights in the indices than the stocks that are still included, it implies that the deletions also have smaller levels of passive ownership than the rest. Thus, if the effects are of this magnitude and modestly significant for the deleted stocks, then it is plausible to think that the effects would perhaps be greater for stocks that have multiple times the passive flows and ownership compared to the deletion sample.

However due to passivity's popularity, by this assertion most of the market would be chronically overvalued which is a dangerous statement and one that is more or less impossible to prove ex-ante. For example, during the tech bubble an investor couldn't argue with a high certainty that the market was overvalued because there were believable fundamentals such as the growth story, that could be used to explain the valuations. Similarly, even if there are chronic overvaluations currently due to passivity, other fundamentals such as low interest rates can be used by market participants to explain stock valuations. Also, recently Jeremy Siegel[35] (2018) noted how the fact that investors can now own diversified portfolios (through index funds) for a fraction of the cost that they used to pay historically, may be a meaningful factor in explaining why market valuations can be higher than before. The idea is that the equilibrium PE ratio for stocks may have been elevated by the amount that was previously paid out to brokers and other intermediaries when investing in stocks. Siegel's notion does seem intuitive and it can be an important reason behind the findings in this thesis and in the research cited previously, but it may also be another seemingly important fundamental that can't necessarily be tested properly just like many of the reasons used by practitioners in the tech bubble. In the end, it is thus only after the correction that we can explain market-wide inefficiencies with a higher confidence and sort out the meaningful fundamentals from the noise.

The results in this paper are nevertheless in line with the previous findings of Morck & Yang (2001) and Belasco et al. (2011) regarding the detachment of indexed stocks and supportive of the decreased price efficiency -findings by Qin and Singal (2015) and the fact that the markets can't be perfectly efficient in any case (Grossman and Stigliz, 1980).

[35] Siegel is a highly decorated professor currently teaching at the Wharton School of Business.

VI – Robustness and limitations

The possibility of reverse causality in the regressions is low due to the nature of index funds and their investors. The possibility that realized skewness or negative deviation could drive the changes in passive ownership is low because cap-weighted funds only buy and sell stocks based on investor flows, i.e. they do not directly react to market movements. Once they replicate the index, they will always follow it approximately 1:1 unless they have investor flows and thus need to sell or buy stocks. Investor flows can of course drive the reverse causality, i.e. make subscriptions and redemptions as a result of market movements, but the flows tend to exhibit opposite patterns than what could explain the results: i.e. positive flows to any type of fund usually coincide with positive market returns (positive skewness) whereas negative flows coincide with negative returns (negative skewness). The result that an increase in passive ownership (positive flows) is associated with negative skewness is thus somewhat contradictory to the classical fund flow and returns chasing -studies and implies that the results are likely not due to reverse causality.

One caveat is however, that for equally weighted index funds, the previously discussed reverse causality would apply: i.e. they own the index's stocks with weights 1/n and if some stocks experience large negative returns, they buy them and sell their other holdings so that the 1/n weight is again reached. For these funds, highly negatively skewed returns would be associated with increased ownership but the causality would be reversed, i.e. skewness would lead to an increase in ownership. Thus, I had excluded these funds from my sample as was mentioned in the data-section. Also, the AS calculations of Cremers helped here for he had not calculated them using equal weight indices. Thus, none of the funds matched with Cremers dataset could be equally weighted, for equal weighted funds would have had a very large active share compared to cap weighted indices. It is still possible that some funds in my sample are equally weighted, but if so then these are certainly in the minority.

Also, as the results also highly indicate that the lagged change of passive ownership is a significant predictor of future realized skewness and negative deviation, it is even more unlikely that there is a reverse causality driving the results, for it would imply that investor flows to passive index funds would somehow be affected by near-future realized tail-movements which tend to be impossible to predict.

The study contains a fair amount of robustness checks in the main results already by having multiple model specifications in each regression. However, the main regressions do not consider the possibility of multicollinearity among the passivity variables affecting the results (as shown by the correlations of the variables in Appendix B), which is why in Table 10 I run the idiosyncratic skewness -regressions separately for each variable. Table 10 further confirms the results in Table 3, indicating that the coefficients are separately as significant in magnitude and statistical significance as in the multivariable setting. Nevertheless, the fact that both the level and the changes are included in the models of Table 3 is important, for it is intuitive that there would be different effects that passive flows can have depending on what a stock's status (i.e. the level of passivity) is in an index, and I thus consider those models to be more representative of passivity's total effects on stock return characteristics.

Finally, various other specifications have been tested outside of this paper but are not shown here due to passivity's consistent effect throughout the results, further confirming the main models discussed in the results.

Table 10: Idiosyncratic skewness regressions with passivity variables individually

The regressions are run for the whole time period of 2000 - 2017 which consists of a total of 284 402 individual firm-quarter findings. Realized skewness is the Fisher-Pearson skewness during each quarter calculated from idiosyncratic returns. P_Own indicates the quarterly ownership-% of all passive funds out of total shares out for each company. ΔP_Own indicates the quarterly percentage point change in the ownership of passive funds for each stock in the sample. ΔP_Own_Lagged is the change during the previous quarter. Russ_Skewness indicates the realized returns distribution descriptor for the Russell 3000 -index during each quarter. P/B indicates a firm-specific valuation metric, calculated each quarter by dividing a firm's market cap with its most recent realized book value of equity. Market cap is the natural logarithm at the quarter end market cap of a stock measured in millions. EPS surprise is the relative surprise of the EPS released in the quarter being analyzed. Lagged variables are calculated from the observations of the previous quarter. Time and sector dummies indicate the amount of dummy variables in the model. The number of observations is not stable in each model for there is lacking data for some variables, most notably the valuation metrics. The t-scores are calculated using robust standard errors.

Dependent variable: Realized Quarterly Idiosyncratic Skewness	Standard OLS regression (eq 5)			Fixed-effects models (eq 6)		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.31	0.31	0.30	0.19	0.19	0.17
	(22.91)	(21.41)	(19.60)	(6.73)	(6.58)	(5.77)
P_Own	-0.006			-0.006		
	(-8.94)			(-6.12)		
ΔP_Own		-0.015			-0.024	
		(-4.48)			(-5.54)	
ΔP_Own_Lagged			-0.010			-0.018
			(-2.45)			(-3.94)
Russ_Skewness	-0.06	-0.06	-0.09	-0.015	-0.014	-0.07
	(-8.22)	(-8.52)	(-10.99)	(-0.27)	(-2.68)	(-1.34)
P/B	0.009	0.009	0.008	0.008	0.007	0.007
	(16.00)	(13.12)	(12.30)	(13.40)	(10.46)	(9.61)
ln(Market cap (m))	-0.03	-0.03	-0.03	-0.02	-0.02	-0.02
	(-14.54)	(-14.99)	(-13.70)	(-11.60)	(-11.48)	(-10.83)
Skewness_Lagged	0.00	-0.001	-0.002	-0.003	-0.006	-0.006
	(0.36)	(-0.54)	(-0.70)	(-1.45)	(-2.32)	(-2.27)
Stdev_Lagged	0.008	0.007	0.008	0.007	0.006	0.008
	(10.99)	(10.26)	(10.66)	(9.64)	(8.63)	(8.99)
EPS Surprise	0.20	0.19	0.18	0.20	0.18	0.18
	(44.30)	(38.83)	(35.19)	(43.95)	(37.56)	(34.96)
Number of time dummies**				71	71	71
Number of sector dummies**				25	25	25
N	163 179	128 826	118 267	163 179	128 826	118 267
Adj. R^2	1.6%	1.5%	1.5%	2.3%	2.3%	2.2%

** Most sector and time dummies are highly significant in all models.

Limitations

There are a few noteworthy limitations in the study. Most importantly, the exact level of passive ownership is impossible to calculate because of missing data on non-U.S. passive institutions, possible data gaps in the CRSP/CDA-structure and unknown investors that mirror passive vehicles through direct investments.

Therefore, the regression results, i.e. “an x-basis point increase in passivity is associated with a y-basis point decrease in realized idiosyncratic skewness” are slightly biased in terms of size, but the direction of the coefficients shouldn’t be affected by this unless the missing sources of passivity data greatly deviate from the U.S. institutions in their trading patterns, which seems rather unlikely. The bias caused by missing data is also likely quite small for the fact that the funds included in the study should cover most of the passive money worldwide.

The study also only covers U.S. listed stocks which implies that there may be some geographic bias driving the results. Therefore, further analysis in other markets such as Europe and Japan would be required. Japan especially could be an interesting candidate for the passivity levels are likely extremely high there due to the Bank of Japan’s monetary easing policies through the purchase of ETFs.

There are also a lot of missing data points on many of the key variables used in the regressions, which is why the sample sizes are not stable across different model specifications, though this limitation’s effect is to some extent diminished by the fact that the sample sizes are still so large in many of the regressions.

Another noteworthy limitation may be a model risk in the sense that a standard linear OLS -setting may not be the mathematically correct way in regressing the third central moment, i.e. the skewness, which may be indicated by the low adjusted R-squared of the regressions, although the methodology used here and hence the low R-squared values does not differ from previous studies around realized skewness (Ben-David et al., 2017). Omitted variable bias is also a slight worry in the skewness regressions for the amount of studies around the subject is so limited to use as a reference point.

Furthermore, the study only analyses the effects of passive ownership and flows but doesn’t have an input on how the active part of the markets interplays with the added passivity. Analysing the flows of active funds and for example the short interest on passivized stocks would be a meaningful addition to the study.

More minor limitations include the various ways of variable computation that have not been tested, such as the way in which stock returns, abnormal returns or skewness-measure can be computed and the different time periods of data that can be used. It is also possible that the corrections made to the passivity and the control variables that were discussed in the data section are insufficient. These issues and their significance has been discussed in more detail earlier however.

VII – Conclusions and topics for further research

In this thesis I study whether the current megatrend of the stock markets, passive investing, affects the characteristics of the stocks that they target and hence the accuracy of stock prices. The intuition behind the study is that any effect caused by passive institutions is non-fundamental because their actions are not based on any firm or market analysis, hence causing decreased price-efficiency.

Additionally, the fact that full or very high levels of passivity would likely lead to market breakdown raises questions: do we observe the negative effects incrementally or do they only appear once a critical threshold has been reached? I study the potential effects in two distinct ways: first by studying the quarterly flows by passive institutions and the changes in ownership at the stock level, where I test whether the flows and level of passive money alter the shape of a stock's realized firm-specific returns distribution and whether there are firm-specific factors enhancing the effects, and second, by studying the potential latent long-term premiums associated with passive money via event study -methodology around index deletions.

In the quarterly flow -analyses I found a statistically significant negative relation of added passivity and the current level of passive ownership on the realized firm-specific skewness of a stock. More specifically, the level of passive ownership decreases the quarterly realized idiosyncratic skewness by about 0.04 for each 10% out of total shares out owned by passive institutions, and that the realized skewness is further decreased by 0.01 – 0.02 for each 1% of shares outstanding purchased by the institutions in the same quarter and by a further 0.01 – 0.02 in the subsequent one

Further analysis through negative deviation -analyses suggests that the *level* of ownership has a different nature in affecting the skewness than the flows of passivity, i.e. the level has a tendency to decrease tail movements overall and hence decrease skewness which supports the previous findings that stocks tend to behave more like the indices in which they are included, whereas the flows have a tendency of increasing left-tail movements, supportive of the hypothesis that passive flows are non-fundamental and may lead to reversals in stock prices. I.e. a 1%-point quarterly increase (flow) in passive ownership leads to an increase of 0.01 – 0.05%-points in realized negative deviation in the same quarter and an additional 0.01 – 0.04%-points in the subsequent quarter and that the level-coefficient predicts a decrease of around 0.03% for 1%-point increase in the level of ownership.

By grouping the sample with each firm's respective ROE-measure, used to proxy for the fundamentals of a company, I found that firms with worse fundamentals tend to have larger reversals as a result of passive money. The results indicate that for firms which do poorly economically, the non-fundamental push of passive flows have a more prominent negative effect on stock price efficiency in the short-term and hence a higher reversal effect once the arbitrageurs correct the errors. The intuition behind this result is that the demand for stocks of poorly performing companies would be low without passive flows and thus the share of non-fundamental demand out of the total demand is higher than what it would be for high performance stocks, thus leading to more prominent mispricings and corrections.

From the event study -analyses I found that stock-deletions that have had more passive ownership in the prior quarter tend to have slightly larger and more significant negative effects as a result of the deletion, implying that passive flows may be affecting the long-term valuations of companies and that the arbitrageurs are limited in their ways of correcting the pricing errors. Furthermore, the most striking difference between more passive and less passive stock-deletions was in the way in which the markets anticipate the deletion, where more passivized stocks have systematically larger reversals in the anticipation period. This indicates that arbitrageurs are willing to take larger bets on stocks that have more passive ownership, which is consistent with the idea that passive flows may be pushing the prices of stocks upwards as the popularity of indexing grows and forces non-fundamental buying pressure. Thus, there is evidence that a long-term "passivity premium" exists.

The results combined suggest that non-fundamental flows induced by the growth in popularity of passive investing can affect the way in which stocks move towards efficient prices and the effects may be quite sticky. Hence, it would seem that added passivity incrementally deteriorates the pricing efficiency for stocks, rather than acts as a binary variable over a certain threshold. For the future, this implies that if the current trend continues and passive vehicles gain large inflows while at the same time active investors face outflows, the potential pricing errors for heavily indexed stocks may increase. This poses an interesting game-theoretical problem caused by indexing: for individual investors it is most favourable to invest in passive funds regardless of the situation (as per Sharpe, 1991) but if everyone does it, market efficiency will break down and ultimately everyone will lose.

The potential topics for future research around passive investing grow in parallel with its popularity: For example, can/do firms use the potential overvaluations in their favour by for example issuing new shares? Is this linked to index additions and deletions? How will the added passivity affect the most important decisions of individual companies that are usually put to shareholder vote? – I.e. as passive institutions target for low tracking error instead of maximizing returns, will they prefer to vote for the less volatile choices, and what happens once their ownership reaches the threshold of 50%, which represents a situation when they can in theory dictate how a company operates. Furthermore, how does the increase in smart beta -fund's popularity link to the popularity of the more simple passive products? For example, if we know that added passivity can alter market efficiency and the way in which stocks behave, then this can potentially create factors in the stock market which factor funds can then follow mechanically (momentum-funds would perhaps be a prime candidate for the analysis), further causing a inefficiencies in the market.

Understanding the implications and effects that added passivity can have on the markets is perhaps one of the most important questions in finance at the moment, for the trend is so strong and the estimates predict extremely high levels of passive ownership already in the near future. The results in this thesis indicate incremental price efficiency deterioration -effects as a result of passivity and it is consistent with the findings in the papers by Morck & Yang (2001), Belasco et al. (2011) and Qin and Singal (2015). Nevertheless, further analysis around the subject is definitely needed to reach a more comprehensive conclusion.

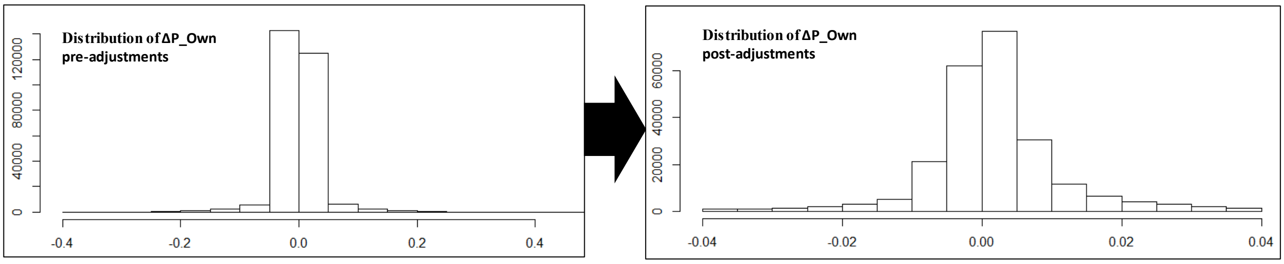
Appendices

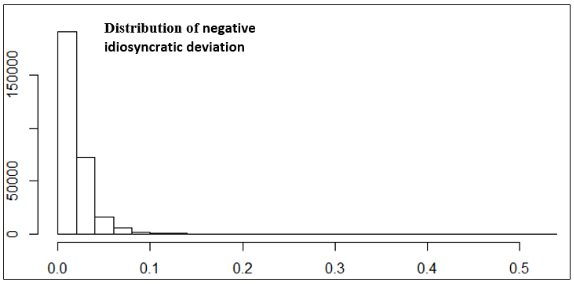
Appendix A – Passive fund -sample

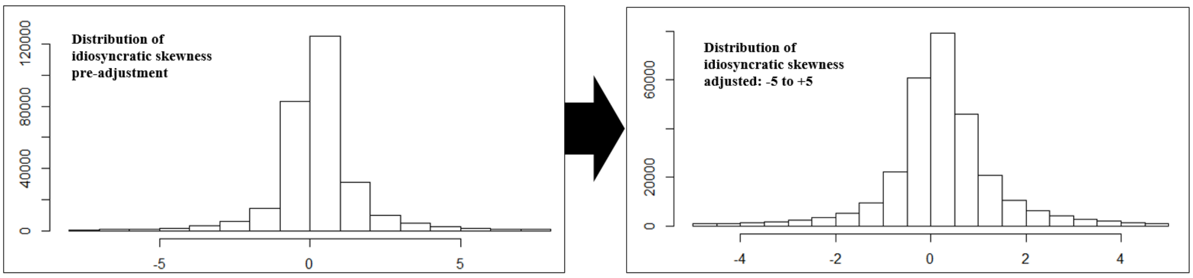
CDA codes for the whole sample of passive funds 2000 - 2017, N=711

391	2994	8850	41073	46106	52877	57032	62164	68323	80428	80837	81141	82564	84899	88707
448	3091	8860	42040	46185	53281	57033	62168	69322	80429	80838	81142	82565	84900	88717
519	3120	9375	42539	46960	53282	57048	62170	69365	80430	80909	81143	82567	84901	89011
676	3169	11050	42543	47050	53617	57049	62171	72729	80431	80944	81161	82569	84902	89038
762	3196	13788	42544	47197	53708	57050	62181	72746	80432	80950	81162	82570	84903	89073
792	3306	14571	42545	47198	53750	57121	62217	72930	80433	80951	81169	82571	84904	89189
805	3307	15058	42547	47199	53751	57124	62219	72986	80434	80952	81173	82573	84905	89190
925	3370	15653	42548	47201	53784	57125	62242	72987	80435	80953	81174	82586	84906	89387
1085	3378	16489	42696	47393	53832	57131	62836	73204	80443	80954	81194	82596	84907	89464
1366	3519	16500	42700	48428	53912	57263	62837	73218	80461	80955	81195	82600	84908	89531
1394	4870	18004	42701	48459	53920	57328	62839	73271	80463	80956	81196	82601	84909	89683
1427	4956	18005	42702	48526	53930	57336	62982	73712	80469	80957	81197	82607	85262	
1586	5105	20408	42703	48638	54018	57337	63069	75704	80547	80958	81198	82617	85344	
1589	5171	20450	42737	48642	54299	57352	63121	75708	80548	80959	81199	82618	86283	
1710	5195	20618	42738	48692	54440	57478	63155	75710	80558	80960	81200	82737	86328	
1903	5378	21325	42739	48755	54460	57589	63603	75711	80559	80962	81201	82861	86330	
1908	5451	21327	42784	49015	54665	57590	63702	75713	80560	80974	81202	82862	86389	
1909	5677	21328	42808	49150	54747	57591	63704	75715	80561	80975	81213	82873	86602	
1927	5705	21859	42809	49197	54843	57592	63836	75716	80562	80977	81304	82881	86883	
1998	5715	21917	42810	49267	55710	57593	63837	75719	80570	81007	81428	82887	87010	
2026	5759	23135	42819	49476	55785	57594	63838	75721	80590	81008	81449	82918	87587	
2035	5794	23138	42921	49645	55811	57595	63839	75722	80591	81011	81454	82945	87588	
2042	5803	23440	42956	50320	55829	57596	64087	75989	80592	81012	81594	82951	87590	
2113	5806	26775	42957	50407	55865	57597	64309	76265	80622	81013	81638	83143	87592	
2152	5941	26782	42974	50440	55965	57612	64346	76581	80624	81014	81639	83154	87593	
2218	5991	26784	42993	50441	56301	57633	64381	77652	80699	81015	81640	83181	87698	
2297	6149	28591	44054	50442	56305	57634	64687	77964	80717	81016	81646	83312	87701	
2326	6200	28898	44055	50443	56342	57829	64725	78284	80718	81017	81669	83329	87703	
2327	6220	28911	44056	50444	56347	58030	64728	78636	80719	81018	81681	83330	87704	
2523	6223	28944	44057	50445	56590	58031	64861	78653	80720	81019	81682	83336	87705	
2547	6232	28984	44058	50446	56611	58032	64863	78665	80721	81020	81683	83341	87706	
2550	6334	29181	44061	50493	56762	58037	64887	78674	80751	81034	81684	83353	87707	
2554	6391	30732	44245	50959	56763	58892	64903	78679	80753	81042	81685	83369	87708	
2559	6485	30797	44276	51131	56769	58893	64918	78687	80754	81044	81686	83377	87709	
2586	6623	30878	44278	51143	56771	58953	64938	78698	80755	81065	81692	83378	87711	
2609	6626	30906	44282	51157	56772	59242	64961	78717	80756	81083	81697	83385	87712	
2627	6627	30907	44283	51528	56773	59243	65260	79197	80757	81084	81783	83395	87713	
2637	6628	32275	44709	51555	56774	59346	66594	79423	80758	81088	81803	83396	87714	
2638	6630	32503	44776	51610	56776	59571	66774	79968	80759	81089	82031	83397	87716	
2648	6631	34218	45073	51619	56790	59622	66840	79970	80760	81091	82219	83401	87717	
2676	6632	35483	45281	51640	56804	59718	66845	79991	80761	81092	82434	83425	87718	
2677	6634	35485	45282	51715	56805	59911	67015	79992	80769	81093	82487	84017	87719	
2705	6750	36450	45404	51801	56807	60026	67040	80205	80770	81094	82528	84210	87746	
2708	6752	36550	45650	51836	56810	60027	67229	80210	80771	81095	82543	84293	87831	
2800	7571	37527	45668	51943	56813	60560	67499	80212	80772	81117	82558	84892	88171	
2839	7598	37696	45669	51951	56871	61412	67519	80213	80773	81121	82559	84894	88218	
2844	7599	37871	45671	51952	56881	61830	67520	80380	80774	81135	82560	84895	88300	
2932	7679	38418	45761	52100	56883	61886	67992	80382	80811	81138	82561	84896	88515	
2934	7995	39977	45921	52150	56906	61943	67996	80396	80824	81139	82562	84897	88519	
2942	8781	39978	45947	52651	57013	62160	67999	80399	80825	81140	82563	84898	88704	

Appendix B – Distributions and correlations of the main variables





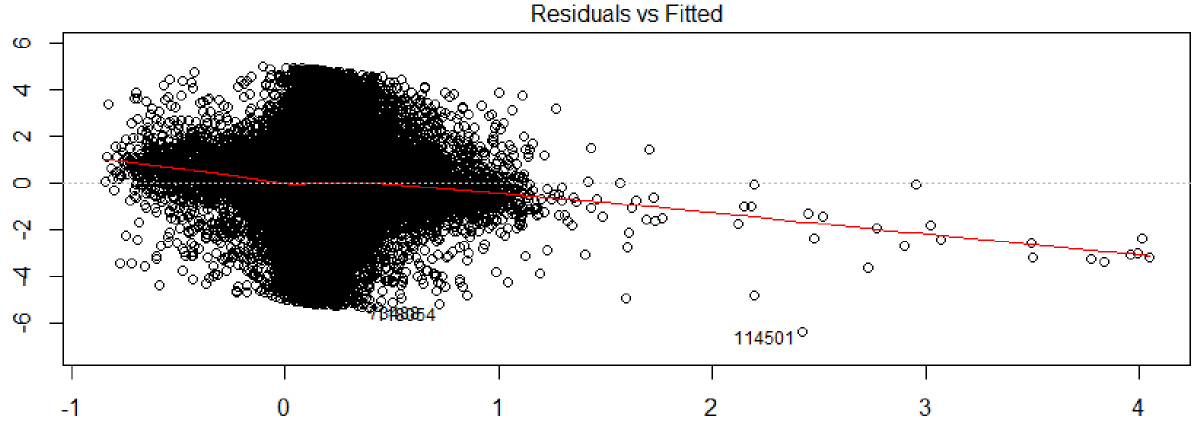


Correlation matrix of the main variables

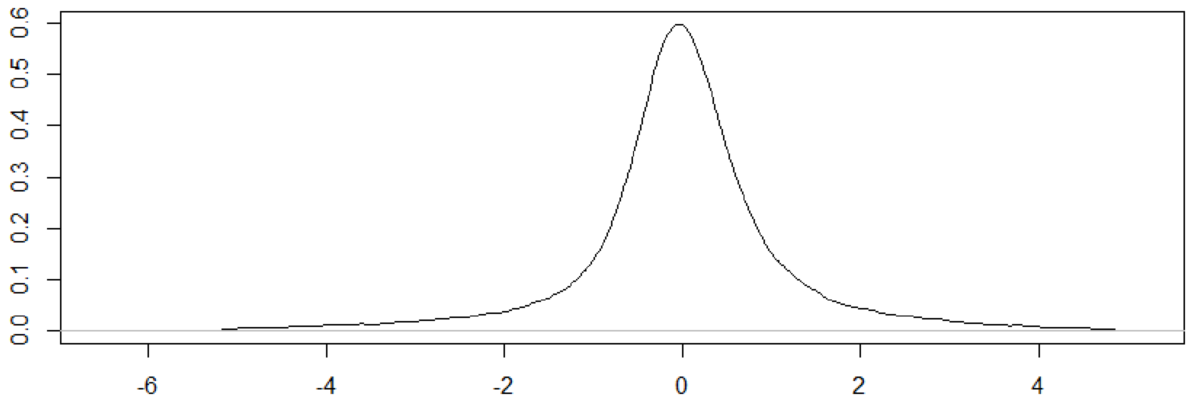
	iret_skew	P_Own	Analysts	PB	Mcap	PS	VOL	Russ_Stdev	Russ_Skew	PE	iret_negdev	Stdev_lagged	Ret_lagged	Skew_lagged	Delta_own_lag	EPS_Surprise	Delta_Own
iret_skew	1	0	-0.03	0.02	-0.01	0.02	-0.02	-0.02	-0.04	0.02	-0.3	0.03	-0.02	0	0	0.13	0
P_Own		1	0.03	-0.02	0	-0.02	0.01	-0.04	-0.13	0	-0.17	-0.11	-0.04	0	0.14	0.01	0.25
Analysts			1	0.12	0.42	0.06	0.45	-0.04	-0.03	-0.01	-0.19	-0.12	-0.04	-0.04	-0.01	-0.04	-0.04
PB				1	0.09	0.16	0.05	-0.06	-0.02	0.08	-0.02	0	0.05	0.01	0	0.01	0
Mcap					1	0.05	0.59	-0.04	-0.03	-0.03	-0.16	-0.11	-0.01	-0.02	-0.02	-0.02	-0.03
PS						1	0.06	-0.05	-0.07	0.26	0.01	0.03	0.04	0	-0.02	0.01	-0.01
VOL							1	0.04	-0.02	-0.01	-0.04	-0.02	-0.03	-0.02	-0.02	-0.01	-0.03
Russ_Stdev								1	0.25	-0.05	0.42	0.25	-0.11	0	0.04	-0.01	0.01
Russ_Skew									1	-0.05	0.2	0.12	-0.09	-0.04	-0.04	-0.01	-0.02
PE										1	0.07	0.07	0.04	0.02	-0.01	0.01	0.01
iret_negdev											1	0.46	0	0.02	0.03	-0.01	0.03
Stdev_lagged												1	0.32	-0.03	0	0.04	0.05
Ret_lagged													1	0.36	0.06	0.08	-0.04
Skew_lagged														1	0.01	0.04	0
Delta_own_lag															1	0.01	-0.16
EPS_Surprise																1	0.01
Delta_Own																	1

Appendix C – Residual plots of main regressions

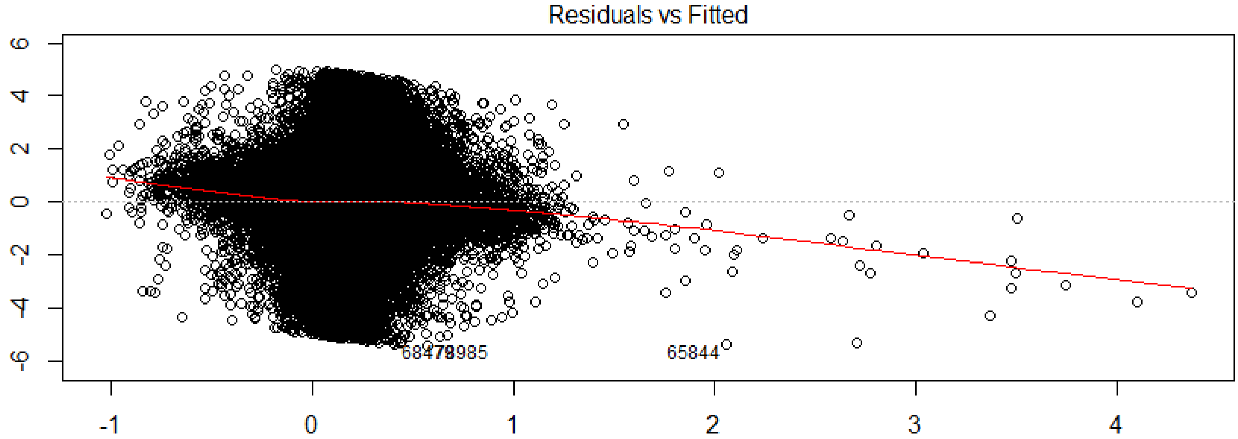
IRS regression Model 3



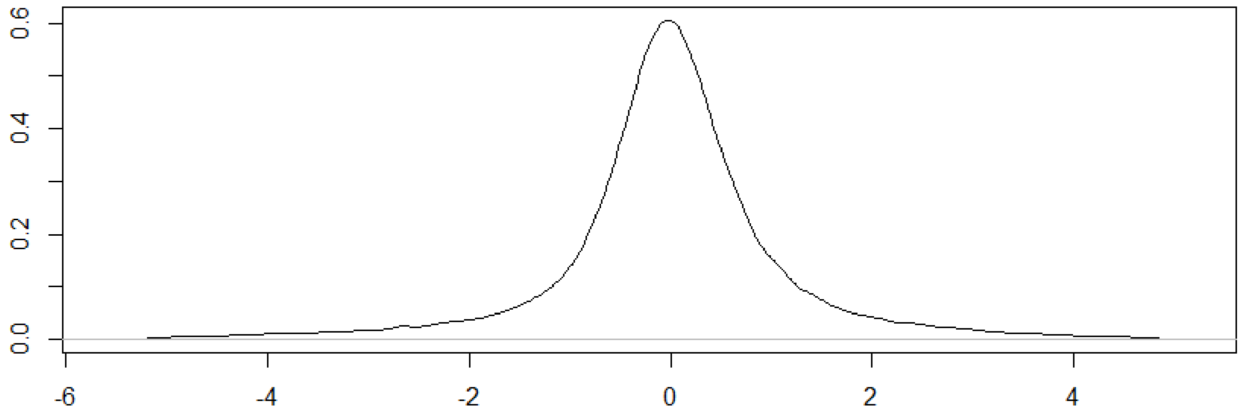
Residual distribution



IRS regression Model 6



Residual distribution



References

Albuquerque, Rui, 2012, “Skewness in Stock Returns: Reconciling the Evidence on Firm Versus Aggregate Returns”, The Review of Financial Studies, Volume 25, Issue 5, Pages 1630–1673.

Arnott Rob, Kalesnik Vitali and Wu Lillian, 2018. “Buy High and Sell Low with Index Funds!”, Research Affiliates -article, June 2018.

Baker, Malcolm, Brendan Bradley, and Jeffrey Wurgler, 2011. "Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly.", Financial Analysts Journal 67, No. 1, Pages 40-54.

Baltussen, Guido, van Bekkum, Sjoerd and Da, Zhi, 2017. “Indexing and Stock Market Serial Dependence Around the World”, Journal of Financial Economics (JFE), Forthcoming.

Banz, Rolf, 1981. "The Relationship Between Return and Market Value of Common Stocks", Journal of Financial Economics 9, Pages 3-18.

Basu, Sanjoy, 1977. “The Investment Performance of Common Stocks in Relation to their Price to Earnings Ratio: A Test of the Efficient Markets Hypothesis”, Journal of Finance 32, Pages 663-682.

Bank of America Merrill Lynch: “The ETF-ization of the S&P 500”, 2018, Equity and Strategy Focus Point, 01/18

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2014. “Do ETFs Increase Volatility?”, Swiss Finance Institute Research Paper No. 11-66.

Black Fischer, 1986. “Noise”, Journal of Finance, Vol. 41, No. 3, Pages 529-543.

Chang, Y. C., H. Hong, and I. Liskovich, 2015. “Regression Discontinuity and the Price Effects of Stock Market Indexing.”, Review of Financial Studies 28, Pages 212-246.

Chen, Hailiang Prabuddha, De Yu Hu, Byoung-Hyoun Hwang, 2014. “Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media”, The Review of Financial Studies, Volume 27, Issue 5, Pages 1367–1403.

Chen, Joseph, Harrison Hong and Jeremy C. Stein, 2000, "Forecasting Crashes: Trading Volume, Past Returns, And Conditional Skewness In Stock Prices," Journal of Financial Economics, 2001, 61, Pages 345-381

Coles Jeffrey, Heath Davidson and Ringgenberg Matthew, 2018, "On Index Investing", *Working paper*

Coval, Joshua, and Erik Stafford, 2007, "Asset Fire Sales (and Purchases) in Equity Markets", Journal of Financial Economics 86(2), Pages 479–512.

Credit Suisse, 2017: "Looking for Easy Games: How Passive Investing Shapes Active Management".

Cremers, Martijn, and Antti Petajisto, 2009. "How active is your fund manager? A new measure that predicts performance.", Review of Financial Studies 22 n. 9, Pages 3329-3365.

D'Avolio, Gene, 2002. "The Market for Borrowing Stock", Journal of Financial Economics, vol. 66, issue 2-3, Pages 271-306

DeBondt, Werner and Richard Thaler, 1985. "Further Evidence On Investor Overreaction and Stock Market Seasonality", Journal of Finance 42, Pages 557-581.

DeBondt, Werner and Richard Thaler, 1987. "Does the Stock Market Overreact?", Journal of Finance 40, Pages 793-805.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990. "Noise trader risk in financial markets", Journal of Political Economy Vol. 98, No. 4, Pages 703-738.

Durnev Artyom, Randall Morck, Bernard Yeung, Paul Zarowin, 2003. "Does Greater Firm-Specific Return Variation Mean More or Less Informed Stock Pricing?", Journal of Accounting Research, Volume 41, Issue 5 Pages 797–836.

Eric, Michael Finke, David Nanigian, 2012. "The impact of passive investing on corporate valuations", Managerial Finance, Vol. 38 Issue: 11, Pages 1067-1084.

Fama, Eugene, 1970. “Efficient capital markets: A review of theory and empirical work”, Journal of Finance 25, 383-417.

French Kenneth, FF3 factors: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Glosten Lawrence, Suresh Nallareddy, and Yuan Zou, 2016. “ETF Trading and Informational Efficiency of Underlying Securities”, Working Paper, Duke University.

Grinblatt Mark, Masulis Ronald and Titman Sheridan, 1984. “The Valuation Effects of Stock Splits and Stock Dividends”, Journal of Financial Economics, Vol 13, Issue 4, Pages 461-490.

Grossman, Sanford and Stiglitz Joseph: “On the Impossibility of Informationally Efficient Markets”, 1980, American Economic Review (70): 393-408.

Israeli, Doron, Charles M. C. Lee, and Suhas Sridharan, 2017. “Is there a Dark Side to Exchange Traded Funds (ETFs)? An Information Perspective”, Review of Accounting Studies, Vol. 22, Issue 3, Pages 1048-1083.

Monteiro, Ana, 2006. “A Quick Guide to Financial Ratios”, The Citizen, Moneyweb Business Insert 6

Moody’s, 2017: “Passive Market Share to Overtake Active in the US No Later than 2024”.

Morck, Randall, F. Yang, 2001. "The Mysterious Growing Value of S&P Index Membership.", NBER Working Paper No. 8654.

Pedersen, Lasse, 2018. “Sharpening the Arithmetic of Active Management”, Financial Analysts Journal, Vol. 74, Pages 21-36.

Petäjistö, Antti, 2011. “The index premium and its hidden cost for index funds”, Journal of Empirical Finance, Volume 18, Issue 2, Pages 271-288.

Petäjistö, Antti, 2013 “Active Share and Mutual Fund Performance”, Financial Analysts Journal, Volume 69, No. 4, Pages 73-93.

Petäjistö, Antti, 2016. “Inefficiencies in the Pricing of Exchange-Traded Funds”, Financial Analysts Journal, Volume 73, No. 1, Pages 24-54.

Qin, Nan and Vijay Singal, 2015. “Indexing and Stock Price Efficiency”, Financial Management, Volume 44, Issue 4 Pages 875–904.

Sapp, Travis and Ashish Tiwari, 2004, “Does Stock Return Momentum Explain the Smart Money Effect?”, Journal of Finance, Volume 59, Issue 6, Pages 2605–2622.

Sharpe William, 1991. “The Arithmetic of Active Management”, Financial Analysts Journal, Vol. 47, No. 1, Pages 7-9.

Shiller, Robert, 1981. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?", American Economic Review 71, Pages 421-436.

Shleifer, Andrei, R. W. Vishny, 1997. "The Limits of Arbitrage.", Journal of Finance 52, Pages 35-56.

Siegel, Jeremy, 2018: “Siegel vs. Shiller: Is the Stock Market Overvalued?”, Knowledge@Wharton, http://knowledge.wharton.upenn.edu/article/siegel-shiller-stock-market/

Staer, Arsenio, 2017. “Fund Flows and Underlying Returns: The Case of ETFs”, International Journal of Business, Vol. 22, No. 4.

Standard & Poors: SPIVA U.S. Scorecard 2016

Sullivan, Rodney N., and James X. Xiong, 2012. “How Index Trading Increases Market Vulnerability”, Financial Analysts Journal 68(2), Pages 70–84.

Thorley, Steven (1999). “The Inefficient Market Argument for Passive Investing”, Unpublished manuscript, Marriot School

Warther Vincent, 1995. “Aggregate mutual fund flows and security returns”, Journal of Financial Economics, Volume 39, Issues 2–3, Pages 209-235.

Wermers, Russ and Tong Yao, 2010. “Active vs. Passive Investing and the Efficiency of Individual Stock Prices”, Working Paper, University of Iowa and University of Maryland.

Wurgler, Jeffrey, 2010. “On the Economic Consequences of Index-Linked Investing.”, NBER Working Paper 16376, National Bureau of Economic Research, Cambridge MA.